


```

QY 412 ctcttcccccggggaagtactgtgacacatgtaccagctcaccagaagccccccgcccac 471
Db 650 ATCTTTCCAAAGTGGACTCTTTCAGGAGACACATTTCGGATTACAAACAGATGCTGG 709
QY 472 acctctgaggaagaaagtgtcgagcccggtgtgtctccctggctagtttagtcc 531
Db 710 GAATTTGGAGAAAGCTTTTGGGGAAGACAAATAGTCCCTTGGAGAGACTTTGGACAG 769
QY 532 ctcttgggacacctggccacctgtggaaaccaggctgcaagccctggccttgcgcccacc 591
Db 770 GCTCTACATGAAGTGCATCCCATCAGTCTTGGGCTGGAGGCCATGGCTCTGAATCCACT 829
QY 592 attgaacctcacctgagcgaggcgacgtgtccattcttctgagtttccacaggttc 651
Db 830 ATTGATCTGACCTGCAATGATTATATTTTCGGTTTGAATTTGACATCTTTACCCGACTC 889
QY 652 tttaagccatggccaaacctcctcaagaactggcagctcctggcagctgcaacacccagc 711
Db 890 TITCAGCCCTGGTCTTTTGTCTCAGGAATGGACAGCTTGTCTGTAATCTCTGCTGGC 949
QY 712 tacatgacctctccatctatgatgaggtccaaagcgtctcgagggcctgagggagcaag 771
Db 950 TACATGGCTTTTTCAGTATGACGAAGTGAAGCTCGCTCCAGAAATTCATTCACAA 1009
QY 772 ccaggcagttacatcttcggcccaagctgtactcctcctggggcagtgggccatcggtat 831
Db 1010 CTGCGCAGTTATATCTTCGGGTGAGCTGACTCTGCTGGGTGAGTGGGCTATTGGGTAT 1069
QY 832 gtgagctcagatggcagcactcctgagaccatcctcctgcaacaaacccctgtcccaggtg 891
Db 1070 GTTACTGTGTGATGGGAATCTTCAGACAAATCCCTCACAATAAACCTCTTCCAGCA 1129
QY 892 cctctggaggagacagaagagcgtctcactctaccagatggaagaccacacaccca 951
Db 1130 CTGATGTGCTTTCAGGGAAGGCTCTATTTGTTTCTGTGATGGACGAATTCAGAACTCT 1189
QY 952 gacctgactgagctggccaggcagacacccagcagccatccacgtgtcagaggagcag 1011
Db 1190 GATCTGTGACTGGCTTATGTGAACCACTCCCAAGACCAATCAAGTACCCAGACAA 1249
QY 1012 ctgagctctactggccatggcactccacatttgagctctgcaagatctgtgtgagc 1071
Db 1250 TAATGAATATATATGTGAGATGGCTCCACATTCACAACTATGTAATAATATGCTGAAAT 1309
QY 1072 aacagagtgatgaagattgagcgtgagggcacctgtctgagctgtcgtgtgtgtg 1131
Db 1310 GATAAGGATGTAAGATGAGCCCTGTGGACACCTCATGTGCATCCTGTCTTACATCC 1369
QY 1132 tggcagcactcggacagcagcactgcccctctgcccgtcgcagatcaaggggctggag 1191
Db 1370 TGGCAGGAATCAGAGGTGAGGCTGTCTTCTGCGGATGTGAATTAAGGTACTGAA 1429
QY 1192 gccgtgagtatctaccagtt 1211
Db 1430 CCCATCGTGGTAGATCCGTT 1449

```

RESULT 2

US-08-822-586-46

; Sequence 46, Application US/08822586

; Patent No. 6015890

; GENERAL INFORMATION:

; APPLICANT: WILLIAM R. JACOBS, JR., JAMES M. MUSSEY AND

; APPLICANT: AMALIO TELENTE

; TITLE OF INVENTION: AN EMBAC OPERON OF MYCOBACTERIA AND

; TITLE OF INVENTION: MUTANTS THEREOF

; NUMBER OF SEQUENCES: 57

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN

; STREET: 90 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

```

; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
; MEDIUM TYPE: DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,586
; FILING DATE: MARCH 20, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/437
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9960
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; US-08-822-586-46

```

Query Match 3.3%; Score 46.4; DB 5; Length 9960;

Best Local Similarity 45.2%; Pred. No. 0.047;

Matches 170; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

```

QY 819 ggccatcggtatgtgagctcagatggcagcatcctcgagaccatccctgccaacaaacc 878
Db 2385 GGGCTCGGGCGCGTGCCTCGCTGGCTTCGGGCGGAACGACATCCCTTCGGATGCTC 2444
QY 879 cctgtccccaggtgctcctggaggacagaagacggtctctacacctaccacagatggaaa 938
Db 2445 CGGGACCCCGCTCATGAGCAGCCCGGACGACAACTTCGGCGACAGCATCCGGCGT 2504
QY 939 gacccacacacacagacctgactgagctggcagcagacacacacacagcagcagcag 998
Db 2505 CGTACCCGACACGAGTTCGGCAGCAGAAAGCGGACACAGCTGCGCGGCGCATCAACGG 2564
QY 999 gtcagaggagagctgagctctactgggcatgagctccacatttgagctctgcaagt 1058
Db 2565 ATCCGCGCGCGCTCGCTAGCGCTGAACCCGCGCACACAGCGCGGTCGCTCGT 2624
QY 1059 ctgctctagagcaacaagagtgatgaagattgagccgtgagggacacctgtctcagctg 1118
Db 2625 GCGTTCGGCGACACAGCAGCCCGCGTCTGCTGCTCGCGTGGTACCGGCTSCCGGACCG 2684
QY 1119 ctgctgtgctgctgctgagcagcactgagcagcagcagcagcagcagcagcagcagcag 1178
Db 2685 CGAACGAGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2744
QY 1179 caagggtcggagggcc 1194
Db 2745 CGAGGTGCAGTGGGCC 2760

```

RESULT 3

US-09-320-878-23/C

; Sequence 23, Application US/09320878A

; Patent No. 6117659

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: BETLACH, Mary C.

; APPLICANT: MCDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 23
; LENGTH: 1693
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-23

Query Match
Best Local Similarity 3.0%; Score 42.2; DB 5; Length 1693;
Best Local Similarity 48.9%; Pred. No. 0.27;
Matches 113; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 797 gctgtactgcctggcgagtgccatcgccatggtatgtgagctcagatggcagcatctgc 856
Db 971 gctgcgaggtccaggtcggtacgcatcgccatcgccatcgccatcgccatcgccatcg 912

QY 857 agaccatccctgcacaaacccctgtccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 915
Db 911 AGACCTTCGACACCCACAGATCGAGACCGAGAGATCGAGACCGATCGAGACCGATCG 852

QY 917 tctacctctaccagatggaagaaacacacacacacacacacacacacacacacacacacac 976
Db 851 TCGACCTTCGCGCGCGCGCGATATCGCGACCTCGACCTCGCTCGCGCGCGATATCGCC 792

QY 977 aaccocagcgcgcatccatcggttcagagcagcagcagcagcagcagcagcagcagcagc 1027
Db 791 AGACCGCGCGCTAGCGGCACCTTCGCGCGCGAGTTCGCGGACCTTCACCTGGG 741

RESULT 4
US-09-320-878-22/c
; Sequence 22, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 22
; LENGTH: 3292
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-22

Query Match
Best Local Similarity 2.9%; Score 41.6; DB 5; Length 3292;
Best Local Similarity 46.3%; Pred. No. 0.45;
Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 73 ctgcagcgcctagaagagcaatcgctgcagcccccgcgggtgtccgtgagtcctccctctgctg 132
Db 3248 CCGCTCGGCCCAATTCGCGCACGCGGTGATACCCGAGCGCTTCGCGGCGCTCCAGGTG 3189

QY 133 cgggacctgtgccccgcacagcgcagctgttcgagagtggtgcccattcttcggcgggcg 192
Db 3188 CGGCGCGATCGCAGGCTGAGGACCTCGCGCGGAGACTCTCGCGCGCGCGGAGCGACC 3129

QY 193 gccggcgaggcgccccgcgggggtccccgggggtcttcgggagcttttactactactactg 252
Db 3128 TTCCGGCGGTGCTCGCCCGCTAGCGGCGGAGAGGTGACCGGTACCGGTAGTCTCGT 3069

QY 253 gccaatctggaggccaaagacagcaggtggccgcgctgtgtgtgtgtgtgtgtgtgtgtgt 312
Db 3058 GAGGGTGTGATGCGCGGGGTGAGGTGCTGCGGAGCTGCTGCGGCGCTCGGTGCG 3009

QY 313 agtgcacacgacgagctcttcctccggcggggtccagactcagggcgacagcagtgccaa 368
Db 3008 CACGGTGAAGAGGTGCCAGACCGGTGCTGCGGCGCGGTACCGCGCAGGCGCA 2953

RESULT 5
US-09-357-072-1
; Sequence 1, Application US/09357072
; Patent No. 6015712
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Brenda F. Baker
; APPLICANT: Hong Zhang
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF FADD EXPRESSION
; FILE REFERENCE: RTS-0027
; CURRENT APPLICATION NUMBER: US/09/357,072
; CURRENT FILING DATE: 1999-07-19
; NUMBER OF SEQ ID NOS: 87
; SEQ ID NO 1
; LENGTH: 1701
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (145)..(771)
US-09-357-072-1

Query Match
Best Local Similarity 2.8%; Score 40; DB 5; Length 1701;
Best Local Similarity 49.1%; Pred. No. 0.88;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 32 agtgggaagaagagcccgccctggcgccggcagtcagagtcgtgcagcgcctagaagagc 91
Db 242 agcgaagctgagcgcgtgcagagcgccctagacctcttcctatgtctgcctgcctgcctg 301

QY 92 aatgcgtcagcccccgctgcgtgagtcctccctcttcgtcgggacctgtgtgcccccca 151
Db 302 acgacctgagcccgccacaccagctcctgcggagctgtcgcctcctcctgcctgcctgc 361

QY 152 cagcgcagctgtcttcagaggggggcccattcttcggcgggcgccggcgagcgcccccg 211
Db 362 acgacctgtgcggcgctgcagcagcttcagagcgggggcgccggcgccggcgccgctg 421

QY 212 ggggtccggcggtctgtgggacttttactatct 247
| | | | | | | | | | | | | | | | | | | | |
Db 422 gggagaagacctgtgtgagcatattaacgtcatat 457

RESULT 6
PCT-US95-16542-1
; Sequence 1, Application PC/TUS9516542
; GENERAL INFORMATION:
; APPLICANT: YEDA RESEARCH AND DEVELOPMENT CO. LTD.
; APPLICANT: WEINURZEL, Henry
; APPLICANT: WALLACH, David
; APPLICANT: BOLDIN, Mark
; APPLICANT: VARFOLOMEY, Eugene
; APPLICANT: METT, Igor
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS/APOL
; TITLE OF INVENTION: RECEPTORS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16542
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 112022
; FILING DATE: 15-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 112692
; FILING DATE: 19-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114615
; FILING DATE: 16-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: WALLACH-16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..768
; PCT-US95-16542-1

Query Match 2.8%; Score 40; DB 6; Length 1701;
Best Local Similarity 49.1%; Pred. No. 0.88;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
QY 32 agtgggaagagggcgccctgtggcgagatcagctgagccttagaagagc 91
| | | | | | | | | | | | | | | | | | | | |
Db 242 AGCGCAAGCTGGAGCGGCTGCAGAGCGGCTAGACCTCTCTCCATCTCTCGTAGCAGA 301
; 92 aatgcgtcgagagggcgccctgtggcgagatcagctgagccttagaagagc 91
| | | | | | | | | | | | | | | | | | | | |
QY 32 agtgggaagagggcgccctgtggcgagatcagctgagccttagaagagc 91
| | | | | | | | | | | | | | | | | | | | |
Db 242 AGCGCAAGCTGGAGCGGCTGCAGAGCGGCTAGACCTCTCTCCATCTCTCGTAGCAGA 301
; 92 aatgcgtcgagagggcgccctgtggcgagatcagctgagccttagaagagc 91
| | | | | | | | | | | | | | | | | | | | |
QY 92 aatgcgtcgagagggcgccctgtggcgagatcagctgagccttagaagagc 151
| | | | | | | | | | | | | | | | | | | | |

Db 302 ACGACCTGGAGCGGCGACACGAGCTCTCTGCGCGAGCTGCTCCCTGCGGCGCC 361
| | | | | | | | | | | | | | | | | | | | |
QY 152 cagcgagctgtctcagagaggtggccattctcggcgggcgagcgagcgagcgccg 211
| | | | | | | | | | | | | | | | | | | | |
Db 362 ACGACCTGCTGGCGCGCTCGACGACTTCGAGGCGGCGCGCGCGCGCGCGCGCTG 421
| | | | | | | | | | | | | | | | | | | | |
QY 212 ggggtccggcggtctgtgggacttttactatct 247
| | | | | | | | | | | | | | | | | | | | |
Db 422 GGAAGAAGACCTGTGTGCAGCAITTAACGTCAAT 457

RESULT 7
PCT-US96-10521-1
; Sequence 1, Application PC/TUS9610521
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
; TITLE OF INVENTION: AND OTHER PROTEINS
; NUMBER OF SEQUENCES: 34
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10521
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,615
; FILING DATE: 16-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,986
; FILING DATE: 17-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 115,319
; FILING DATE: 14-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 116,588
; FILING DATE: 27-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 117,932
; FILING DATE: 16-APR-1996
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..768
; PCT-US96-10521-1

Query Match 2.8%; Score 40; DB 6; Length 1701;
Best Local Similarity 49.1%; Pred. No. 0.88;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
QY 32 agtgggaagagggcgccctgtggcgagatcagctgagccttagaagagc 91
| | | | | | | | | | | | | | | | | | | | |
Db 242 AGCGCAAGCTGGAGCGGCTGCAGAGCGGCTAGACCTCTCTCCATCTCTCGTAGCAGA 301
; 92 aatgcgtcgagagggcgccctgtggcgagatcagctgagccttagaagagc 151
| | | | | | | | | | | | | | | | | | | | |
Db 302 ACGACCTGGAGCGGCGACACCGAGCTCTCGCGAGCTCTCGCTCCCTCGCGGCGCC 361
| | | | | | | | | | | | | | | | | | | | |
QY 152 cagcgagctctcagagaggtggccattctcggcgggcgagcgagcgagcgccg 211
| | | | | | | | | | | | | | | | | | | | |
Db 362 ACGACCTGCTGGCGCGGTGACGACITCGAGGCGGCGGCGGCGCGCGCTG 421

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/672,211
FILING DATE: 27-JUN-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4692 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-672-211-1

Query Match 2.7%; Score 38.4; DB 3; Length 4692;
Best Local Similarity 50.9%; Pred. No. 2.8;
Matches 116; Conservative 0; Mismatches 111; Indels 1; Gaps 1;
QY 4 gctctggcgtgcccgcgtggggcgcacagtggaagagccgcgccttggccgggca 53
DB 1833 GCTTGGGCTGGGACGACGACGGGAG-CCTTGGGCCCGGACGGGGCGACGGGTGCGGGCTG 1775
QY 64 gtacgatgctgcagcgcctagagagcaatgcgtcgacccccgcgtgcccgtgagtc 123
DB 1774 TCCAGGGGCTGCAGAGGCTGCCGGCCGACGAGGGCGGCGCTGTCGGGGCTGCC 1715
QY 124 ccttcgtcgggacctgctgccccgcacagcagcactgcttcagagtggtgcccattct 183
DB 1714 TCCAGGCTGCGGAGTCCCTCCCGACGACCATGTGCTGTGCTTGAAGAGTTATAG 1655
QY 184 cggcggcggccggcggagcggcccggggtcccgggggtctctggg 231
DB 1654 ATCTGTGTAACCTTCGCGCGGCGGCGGTGCGCCCGGCGCTCGTG 1607

RESULT 13
US-09-225-170-1/c
Sequence 1, Application US/09225170
Patent No. 6017763
GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip Thomas
APPLICANT: Braselmann, Sylvia
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds, LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: Fastseq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/225,170
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/916,917
FILING DATE: 15-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Abrams, Samuel B
REGISTRATION NUMBER: 30,605
REFERENCE/DOCKET NUMBER: 8549-0006-999
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4692 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-225-170-1

Query Match 2.7%; Score 38.4; DB 5; Length 4692;
Best Local Similarity 50.9%; Pred. No. 2.8;
Matches 116; Conservative 0; Mismatches 111; Indels 1; Gaps 1;
QY 4 gctctggcgtgcccgcgtggggcgcacagtggaagagccgcgccttggccgggca 63
DB 1833 GCTTGGGCTGGGACGACGACGGGAG-CCTTGGGCCCGGACGGGGCGACGGGTGCGGGCTG 1775
QY 64 gtacgatgctgcagcgcctagagagcaatgcgtcgacccccgcgtgcccgtgagtc 123
DB 1774 TCCAGGGGCTGCAGAGGCTGCCGGCCGACGAGGGCGGCGCTGTCGGGGCTGCC 1715
QY 124 ccttcgtcgggacctgctgccccgcacagcagcactgcttcagagtggtgcccattct 183
DB 1714 TCCAGGCTGCGGAGTCCCTCCCGACGACCATGTGCTGTGCTTGAAGAGTTATAG 1655
QY 184 cggcggcggccggcggagcggcccggggtcccgggggtctctggg 231
DB 1654 ATCTGTGTAACCTTCGCGCGGCGGCGGTGCGCCCGGCGCTCGTG 1607

RESULT 14
US-08-680-506-5/c
Sequence 5, Application US/08680506C
Patent No. 6008013
GENERAL INFORMATION:
APPLICANT: Reynolds, Paul R.
TITLE OF INVENTION: CHONDROCYTE PROTEINS
FILE REFERENCE: 176/60091
CURRENT APPLICATION NUMBER: US/08/680,506C
CURRENT FILING DATE: 1996-07-08
EARLIER APPLICATION NUMBER: 60/021,672
EARLIER FILING DATE: 1996-07-05
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 5
LENGTH: 954
TYPE: DNA
ORGANISM: Gallus gallus
US-08-680-506-5

Query Match 2.7%; Score 38.2; DB 5; Length 954;
Best Local Similarity 50.2%; Pred. No. 2;
Matches 120; Conservative 0; Mismatches 118; Indels 1; Gaps 1;
QY 103 ccccggtctcgtgagtcctcccttcgttggggacctggtgccccgcacagcgagctg 162
DB 390 CGCCAGGTGTTCCATGTCGGGCGGCGCGCGCCGACGCGCGCTCGAGGCTCTCCTC 331
QY 163 ctttcagaggtggccattcttcggcgagcggcgagggcccgcccggggtcccgcc 222
DB 330 GTACCACATCGGCCAGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 271

QY	223	ggctctgggaactttctactctactctccctggccaatctg-gaggccaaagacagcagcagt	281
Db	270	gcggagcggccctgcagctgcgccaggcggcggaattgcccgccgctggsgtatcgccgt	211
QY	282	ggccgcgcgtgctccctccccggggccgaagagatgccaacgcagagctcttcggggcgg	340
Db	210	gcgctgcgcgaacacacgcggcgagctgtcagcggcgccaggcggcgggcagcggcgcg	152

RESULT 15

```

US-08-680-506-8/c
; Sequence 8, Application US/08650506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROLIFERATION
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680-506-8
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,611
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-8

```

Query Match 2.7%; Score 38.2; DB 5; Length 1347;
Best Local Similarity 50.2%; Pred. No. 2.2;
Matches 120; Conservative 0; Mismatches 118; Indels 1; Gaps 1;

Search completed: December 7, 2000, 13:40:43
Job time: 27492 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 18:06:44 ; Search time 152.48 Seconds
(without alignments)
3503.360 Million cell updates/sec

Title: US-09-434-708-3

Perfect score: 1422
Sequence: 1 atggcttgcggtggccccc.....ccacgacccctgccccggcc 1422

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 18783133 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36.*

1: /SID86/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID86/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID86/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID86/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID86/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID86/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID86/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID86/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID86/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID86/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID86/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID86/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID86/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID86/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID86/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID86/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID86/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID86/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID86/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID86/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID86/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	340	23.9	3090	20 X89079	Nucleotide sequenc
2	48	3.4	10732	21 A10594	Gene encoding a su
3	46.4	3.3	9960	19 V58939	Mycobacterium smeg
4	46	3.2	114955	20 X53491	Human adenosine Al
5	42.2	3.0	1693	21 Z56005	Contig 004 from co
6	41.6	2.9	1140	21 Z87290	S. venezuelae deso
7	41.6	2.9	3291	21 Z56004	Contig 003 from co
8	41.6	2.9	12441	21 Z87284	S. venezuelae deso
9	41.6	2.9	13613	21 Z87319	S. venezuelae deso
10	41.4	2.9	1582	20 X08910	Human FADD protein
11	41.4	2.9	1642	17 T39397	FADD (Fas-associat
12	41.2	2.9	1170	20 Z39042	Human fishboy G-pr

C 13	41.2	2.9	114955	20 X53491	Human adenosine Al
C 14	40.4	2.8	880	17 T38913	B2-bradykinin rece
15	40	2.8	1701	17 T30372	MORT-1 cDNA. Homo
16	40	2.8	1701	18 T61397	MORT-1 coding sequ
17	40	2.8	1701	21 Z44745	Human FADD DNA. H
18	40	2.8	53789	19 V21187	Amcyotatopsis medi
19	39.8	2.8	627	20 V71930	MORT1 isoform MORT
C 20	39.2	2.8	1308	17 T31602	Cartilage-derived
C 21	39.2	2.8	2170	21 Z98402	Porcine betal-adre
C 22	39.2	2.8	15672	12 Q10613	Rianodin receptor
C 23	38.8	2.7	738	20 X86968	PHA acetateacetyl-Co
C 24	38.8	2.7	3632	19 V61586	Alpha-1A calcium c
25	38.8	2.7	6225	20 X52773	Human enzyme-relat
26	38.8	2.7	6225	21 A34721	Human adenosine re
C 27	38.8	2.7	6436	20 X86965	DNA containing PHA
C 28	38.8	2.7	77536	21 A14651	Nucleotide sequenc
C 29	38.4	2.7	1746	11 Q05558	Sequence encoding
C 30	38.4	2.7	2914	20 X57059	Human KCNO3 cDNA.
C 31	38.4	2.7	4692	19 V04633	Porcine phosphoino
C 32	38.4	2.7	4692	20 V74099	Porcine G-protein
C 33	38.4	2.7	4692	21 Z88812	Pig p10i regulator
C 34	38.2	2.7	954	19 V19862	Gallus domesticus
C 35	38.2	2.7	1347	19 V19864	Gallus domesticus
C 36	38.2	2.7	1587	19 V19863	Gallus domesticus
C 37	38.2	2.7	2233	19 V19861	Gallus domesticus
C 38	38.2	2.7	5027	19 V19860	Gallus domesticus
C 39	38	2.7	1959	11 Q04784	Sequence encoding
C 40	38	2.7	2150	20 X87820	Mouse N-acetylgluc
C 41	38	2.7	3318	15 Q68194	Ins5099-10 transpo
C 42	38	2.7	5392	15 Q64201	Sequence comprisin
C 43	38	2.7	15377	13 Q25975	MH mutant porcine
C 44	38	2.7	44377	18 T78508	Platenolide synth
C 45	38	2.7	44377	18 T80414	Platenolide synth

ALIGNMENTS

RESULT 1

X89079
ID X89079 standard; DNA; 3090 BP.
XX
AC X89079;
XX
DT 14-SEP-1999 (first entry)
XX
DE Nucleotide sequence of human Cbl.
XX
KW LAT; tyrosine kinase; linker for activation of T cell; TCR; human;
KW T-cell receptor; TCR signalling pathway; neoplasia; inflammation;
KW hypersensitivity; allergy; microbial infection; genetic disease;
KW autoimmune disease; graft rejection; modulator; Cbl; ss.
XX
OS Homo sapiens.
XX
PN WO9932627-A2.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98WO-US27400.
XX
PR 23-DEC-1997; 97US-0058690.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Samelson LE, Zhang W;
XX
XX WPI; 1999-418925/35.
DR -P-PSDB; Y27126.
XX
XX Linker for activation of T cell protein used to, e.g. screen for
modulators of T cell signalling

Disclosure; Fig 12A; 125pp; English.

The invention relates to a protein tyrosine kinase substrate LAT (linker) for activation of T cells) protein. Modulation of interaction between LAT and the T-cell receptor (TCR) affects the TCR signalling pathway. LAT is a substrate for tyrosine kinases and becomes phosphorylated after TCR engagement, resulting in recruitment of other signalling molecules. LAT is used to identify and test (antagonists of tyrosine kinase signalling pathways, i.e. modulation of interaction between tyrosine kinase substrates and intracellular ligands or between these ligands and other members of the pathway, including identification of downstream signalling proteins, particularly in immune system cells. These modulators are potentially useful as drugs and diagnostic agents, particularly for diseases that involve undesirable cell proliferation, differentiation, growth or T cell energy, e.g. neoplasia, inflammation, hypersensitivity/allergy, microbial infection, metabolic, genetic or autoimmune diseases, graft rejection. LAT is also used to generate specific antibodies, used for detection of LAT. Nucleic acid that encodes LAT, or its fragments, are used to identify homologous sequences in other species; to detect the LAT gene and as sources of antisense therapeutics. Modulators of LAT are potentially more specific and less toxic than known immunosuppressants such as cyclosporin. The present sequence represents the nucleotide sequence of human Cbl.

Sequence 3090 BP; 757 A; 840 C; 764 G; 729 T; 0 other;

```
Query Match      23.9%; Score 340; DB 20; Length 3090;
Best Local Similarity 62.2%; Pred. No. 3.3e-64;
Matches 535; Conservative 0; Mismatches 325; Indels 0
```

352	QY	agcgacagctggccaaagctggccatcatcttcacccacatgcacgagagctgcacgca	411
590	Db		
590	agcgaaacctaaacaaactgctccctcatcttcagccacatgctggcagaactaaaggga	649	
412	QY	ctctcccggggcaaaagctactgtgagacatgtaccagctccaacaaggccccccac	471
650	atcttccaagtgactctttcaggagacacatttcggtattactaaagcagatgctgcg	709	
472	QY	acctttgaggggaaagtctcgagagcccggtgtgtgcctcgctgggtgagttgagtc	531
710	gaattttgagaaaagcttttgggaaaaagacaatgctccttggaaagcttttcgacg	769	
532	QY	ctctcgggcacctgcacccctgtgaaacagagctgcacagccctggctgtgcaccac	591
770	gctctacatgaagtgcattcccatcagttcttgggtcggaggccaaggctcgaatcccat	829	
592	QY	attgaacctcacctgcagcgggcagctgtccattcttcgagttcgacgtcttcacagctc	651
830	attgacttgacgtcgaatgattatttcggttttgaaattgacatcttttaoccgact	889	
652	QY	tttcagccatgcccacaactctcgaagactggcagctctcgttcgagtcacaaccacagc	711
890	tttcagccctggctccttctgcagggaattgggaacagcctgtgtcgaetcatctcggc	949	
712	QY	tacatggccttctcacctatgatgagtcacaagagctctgcaggcctgcagggaacag	771
950	tacatggctttttgacgtatcacgaagtgaagctgcgtcccgagaattcattcacaaa	1009	
772	QY	ccaggcagttacatcttcggcccgacgtgactgcctcgggcagtgaggccaatggctat	831
1010	cctggcagttatatttcggctgagctgtactcgtcgggccaagtgggtattgggtat	1069	
832	QY	gtgagctcagatgycagacatcctgcagaccatccctgcacaacccctgtcccaggtg	891
1070	gttactgctgtaggaaacattctcagacaaatccctcacaataaacctctcttccaagca	1129	
892	QY	ctcctggaggccagaagagcgcttctacctctaccagatgaagaaagccacaacca	951
1130	crgattgagcgttcagggaagccttcatttgtttcctgatgagcaaacagaatcct	1189	
952	QY	gacctgactgactcggccaggcagaacccacagcagcgatccactcgtgtcagaggacag	1011

QY	377	tcctcttcagccacatgcacgcagatgcacgcacactctcccccggggccaaagtactgtg	436
Db	9814	thrsygyrasashasysvaaarghsvaaenashaashaasraraaggyrvaahs	9873
QY	437	gcacatgtaccagctcacaagggcccccacacacctctggaggaaagtgcggag	496
Db	9874	asthrgnasrnrgncysncysrassrsvavaasnraraaggyaanaasr	9933
QY	497	ccgggtgtgtgcctcctggctgagtttgagtcctcctcctggcactgccacctgtgg	556
Db	9934	sargrsrgnassrsgysgvaargraavaastyraassrarsgnygygyarg	9993
QY	557	aaccagctgcacagccctggctgcgcacacccattgaactcactgcagcgggcag	616
Db	9994	ashrtyrasagymtrthrgytrmtvaargvahnurysrsgrysvagghsv	10053
QY	617	tgccatcttcagttcacgtcttcaccaggtctttccagccatggccacactctcca	676
Db	10054	acystyrasryvsthrasyryvaahnaagythrraastrmtashscyshsrgn	10113
QY	677	agaactggcagctcctggcagtcacacccacccaggtacatggcctctcactatgatg	736
Db	10114	asvasrmtaragaasysygnhastrasgthrgrasgyggaasngthrhsvagysg	10173
QY	737	agctccacagcgtctgcagcctgcaggacacagccagcagcttacatctccgcacca	796
Db	10174	argnsraragasnrsrgrnvatnrtrscysgyhsaasnaasngytrtyrasngysg	10233
QY	797	gctgtactcctggcggcagtgccatggcctatgtgagctcagatggcagcatcctgc	856
Db	10234	hthrasysasysraagrgnysgvagarysgnsraathrasarggsrmtasngnngs	10293
QY	857	agaccatcctgccaaacacccctgtcccaggtgtcctggagtcagacagaagcgct	916
Db	10294	rgysthrasrrarngnsrnysgsrasngyngasngysasngysasnsrasnsgnh	10353
QY	917	tcctcctaccagatggaagaccacacccagcctgactgagctcgccaggcag	976
Db	10354	sthrgaagysaaaahasnysvhsrgggaasnyasnsrasnysggaaaycysgngn	10413
QY	977	aaccacagcgcctcacgtgtcagagcagcagctgcactctactgggcatggact	1036
Db	10414	gsrmtsgycysrshasasysaragsnasnastgrthraaasrgysaagcysg	10473
QY	1037	ccacattgagctcgaagatctgctgagcaacaaggatgtgaagattgagccgt	1096
Db	10474	ngthrasnyysgnysaamttaaarysasaasrhasrvaaathrgnhaenthrvath	10533
QY	1097	gcgggcacgtctgcagctgcctgcctgcctggcagcactcggacagccagact	1156
Db	10534	rasnaenthratthrtrthraagaasnaasrrhsrsvaysvayathragsr	10593
QY	1157	gccccttcgcgcgcagatcaaggctgggagcgctgagtctatccaggtccacg	1216
Db	10594	asgmmtaasasrsgnrgysaasrasgysasnsrasnrgyvagrgysvaasnyvays	10653
QY	1217	gtcagctactcctgagcactcagggaacagcagctgaccaggagcgaggttgagac	1276
Db	10654	gyhsasargasnthraasasnaavaarysrgysrsgystrargysaigaigysys	10713
QY	1277	tgsggagcag	1285
Db	10714	arysraasr	10722
RESULT	3		
V58939			
ID	V58939	standard; DNA; 9960 BP.	
XX	AC		
XX	V58939;		
XX	15-FBB-1999	(first entry)	

XX Mycobacterium smegmatis embCAB operon.
DE
XX
XX Drug resistance; antibiotic resistance; antimycobacterial;
KW ethambutol; embCAB operon; infection; vaccine; therapy; ds.
XX
XX Mycobacterium smegmatis.
OS
XX WO9841533-A1.
PN
XX 24-SEP-1998.
PD
XX 16-MAR-1998; 98WO-US05128.
PF
XX 20-MAR-1997; 97US-0822586.
PR
XX {YESH } UNIV YESHIVA EINSTEIN COLLEGE.
PA
XX Jacobs WR, Musser J, Telenti A;
PI
XX WPI; 1998-521160/44.
DR
XX P-PSDB; W73053-57.
DR
XX
XX Wild type and mutated sequences of Mycobacterium embCAB operon -
PT useful to e.g., identify ethambutol-resistant mycobacterial strains
PT and produce antisense sequences to treat mycobacterial infections
PT
XX
XX Disclosure; Fig 6A-I; 62pp; English.
PS
XX
XX This is the DNA sequence of the Mycobacterium smegmatis embCAB
CC operon, which determines resistance to the antimycobacterial drug
CC ethambutol (EMB). It includes the emba, embc and embb genes that
CC encode proteins (see W73053-54) which are the target of action of
CC M. smegmatis for EMB. To identify genes conferring EMB resistance,
CC a genomic library from a high level EMB-resistant mutant of M. Four
CC smegmatis was introduced into wild-type M. smegmatis mc2155. Four
CC overlapping cosmids were identified which conferred a resistant
CC phenotype. The minimum size fragment capable of conferring EMB
CC resistance was 9 kb (pIMW99). pIMW99 plus 7 kb upstream M.
CC smegmatis sequence was sequences revealing 3 homologous open
CC reading frames (embC, embB, embB) and 4 additional potential coding
CC regions. Wild-type and mutated embCAB nucleic acid sequences are
CC useful as probes used in the diagnosis of drug-resistant
CC mycobacteria or to determine the susceptibility of mycobacteria to
CC EMB. The nucleic acids are also useful in the treatment of
CC mycobacterial infections; anti-DNA or anti-RNA sequences can be
CC administered to inhibit embCAB operon mRNA activity (claimed). The
CC invention additionally provides for the use of embCAB operon
CC nucleic acid sequences as vaccines, or to improve existing
CC vaccines.
XX
XX Sequence 9960 BP; 1389 A; 3592 C; 3360 G; 1618 T; 1 other;

```
Query Match          3.3%; Score 46.4; DB 19; Length 9960;
Best Local Similarity 45.2%; Pred. NO. 0.28;
Matches 170; Conservative 0; Mismatches 206; Indels 0; Gaps 0;
```

[illegible][illegible]

KW picromycin; antibiotic production; narbomycin; ds.

XX OS	Streptomyces venezuelae.
XX FH	Location/Qualifiers
FT CDS	/complement [694..1692]
FT FT	*tag= d
FT FT	/product= SAM_synthase
FT FT	/note= "Partial S-adenosylmethionine synthase"
FT FT	/complement [3..692]
FT FT	*tag= b
FT FT	/product= ORF16_product
FT FT	/note= "M. tuberculosis cbhk homologous partial protein"
PX PN	WO9961599-A2.
XX XX	
PD PD	02-DEC-1999.
PF PF	27-MAY-1999; 99MO-US11814.
PR PR	28-MAY-1998; 98US-0087080.
PR PR	28-AUG-1998; 98US-0141908.
PR PR	22-SEP-1998; 98US-0100880.
PR PR	08-FEB-1999; 99US-0119139.
PA PA	(KOSA-) KOSAN BIOSCIENCES INC.
PI PI	Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
DR DR	WPI; 2000-072618/O6.
DD DD	P-PSDB; Y67216, Y67217.
XX PT	New recombinant DNA encoding a domain of narbonalide polyketide synthase, for production of ketolide antibiotics -
PP PS	Disclosure; Page 38; 98pp; English.
CC CC	This is contig 004 from the recombinant cosmid pKSO5023-27 DNA sequence (see Z56001) which contains a Streptomyces venezuelae DNA insert. The cosmid contains open reading frames which encode the various modules of the narbonalide polyketide synthase (PKS). The invention relates to recombinant DNA containing a coding sequence for a narbonalide PKS. Polyketides are compounds synthesised from 2-carbon units through a series of condensations and subsequent modifications. Modular PKSS are responsible for the production of many antibiotics including piromycin. The narbonalide PKS consists of a loading module, six extender modules, and two thioester domains. Four proteins make up the narbonalide PKS (PICAI, PICAILI, PICAIII and PICALV). PICALI includes the loading module and extender modules 1 and 2, PICALII includes extender modules 3 and 4, PICAIII includes extender module 5 and PICALV includes extender module 6 and a type II thioesterase domain. The second type II thioesterase domain is found on the PCB protein. The nucleotide sequences encoding all of these proteins can be isolated in recombinant form from the recombinant cosmid pKSO5023-27. Narbonalide is desaminolyated in S. venezuelae to yield narbomycin, the desoaminyl transferase enzyme is required for this conversion, and the desamine biosynthetic genes are also found in cosmid pKSO5023-27. The recombinant DNA of the invention is used to express, in transformed cells, narbonalide (or its derivatives) or other ketolides (particularly hybrids), which may then be converted (e.g. by other enzymes recombinantly expressed in the same hosts) to polyketide antibiotics or their intermediates. The antibiotics are useful in human or veterinary medicine.
XX SQ	Sequence 1693 BP; 237 A; 566 C; 633 G; 257 T; 0 other;

Query Match	3.0%; Score 42.2; DB 21; Length 1693;
Best Local Similarity	48.9%; Pred. No. 1,6;
Matches 113:	Conservative 0; Mismatches 118; Indels 0; Gaps

```
Qy      797 gctgtactcgcctgggscagtgcccatcgctatgtgatctcacattgccacatcttcg 856
       ||||| ++++++ |||||| | ||||| ++++++ ||||| ++++++ |||||
bb     971 GCTCGAGGTCACAGTGCCTGTACGGATCGCAAGGCCGAGCCTGCGCTTTCTGCG 912
        ++++++ ++++++ ++++++ ++++++
```

CC engineer PHA monomer synthases or to prepare biologically active agents,
 CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
 CC chronic obstructive pulmonary disease as well as other diseases involving
 CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
 CC antibiotics which are active against a variety of organisms, e.g.,
 CC bacteria, including multi-drug resistant pneumococci and other
 CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
 CC protection agents (e.g., fungicides or insecticides) via expression of
 CC polyketides in plants. Sequences 287286-287294 represent desosamine
 CC biosynthetic genes from Streptomyces venezuelae ATCC 15439, which encode
 CC proteins Y77181-Y77189.
 XX
 SQ

Sequence 1140 BP; 163 A; 472 C; 373 G; 132 T; 0 other;

Query Match 2.9%; Score 41.6; DB 21; Length 1140;
 Best Local Similarity 46.3%; Pred. No. 2.1;
 Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 73 ctacagcgctagaagcaatcgctgaccccccgcgtgtccgtgagtcctcccttcgctg 132
 DB 1125 CGCTCGGCCCATTCGCGACGGCTCGATCACCGCAGCGCTCGGGCGGTCCAGGTG 1066
 QY 133 cgggacctgctgcccgcacagcgagctgtcttcgagaggtggcccatcttcggggggc 192
 DB 1065 CGGGCCGATCGGCAGGCTGAGGACCTGCCGCGGAAGCTCTCGGCCCGCGGAGCGAGCC 1006
 QY 193 gcggcgagagggcccccgggggtccgcgggtcttgggaactttctactatctacctg 252
 DB 1005 TTCGGCGGTCCTCCCGCGGTAGCGCGGCGAGAGGTGCACGGGTACCGGTAGTGCCT 946
 QY 253 gccaatctgagggcgaagcagcgaggtggcgcgctgctcctcccgggcggaag 312
 DB 945 GAGGGTGTGATCCCGGGGCTCGAGGTGGCTGGCACCTGTGTCGGCGGCTCGGTGCG 886
 QY 313 agtgcacagcagcttcttcggggggggtccagactcagcagcagctggccaa 368
 DB 885 CACGGTGAAGAGGTGCACACCGGGTCGGTGTGCGGCGCGGTACCGGAGGCGCA 830

RESULT 7

256004/C
 ID 256004 standard; DNA; 3291 BP.

XX AC 256004;

XX DT 23-MAR-2000 (first entry)

DE Contig 003 from cosmid pKOS023-27 from Streptomyces venezuelae.

XX Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; contig 003;
 KW ketolide; dNDP glucose synthase; dehydratase; picromycin; narbomycin;
 KW antibiotic production; ds.
 XX

OS Streptomyces venezuelae.

XX Key Location/Qualifiers

FT CDS 104..982

FT /*tag= a

FT /product= dNDP_glucose_synthase

FT /note= "glucose-1-phosphate thymidyl transferase"

FT CDS 1114..2127

FT /*tag= b

FT /product= dNDP_glucose4_6_dehydratase

FT 2124..3263

FT /*tag= c

FT /product= PICCI

FT /transl_except= (Pos:2874..2876, aa:Xaa)

FT /note= "Xaa = Unknown"

XX WO961599-A2.

XX PD 02-DEC-1999.

XX 27-MAY-1999; 99WO-US11814.
 XX
 XX 28-MAY-1998; 98US-0087080.
 XX 28-AUG-1998; 98US-0141908.
 XX 22-SEP-1998; 98US-0100880.
 XX 08-FEB-1999; 99US-0119139.
 XX (KOSA-) KOSAN BIOSCIENCES INC.

PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;

XX WPI; 2000-072618/06.

DR P-PSDB; Y67206, Y67214, Y67215.

XX New recombinant DNA encoding a domain of narbonolide polyketide
 PT synthase, for production of ketolide antibiotics -

XX Disclosure; Page 37-38; 98pp; English.

XX This is contig 003 from the recombinant cosmid pKOS023-27 DNA sequence
 CC (see 256001) which contains a Streptomyces venezuelae DNA insert. The
 CC cosmid contains open reading frames which encode the various modules of
 CC the narbonolide polyketide synthase (PKS). The invention relates to
 CC recombinant DNA containing a coding sequence for a narbonolide PKS.
 CC Polyketides are compounds synthesised from 2-carbon units through a
 CC series of condensations and subsequent modifications. Modular PKSs are
 CC responsible for the production of many antibiotics including picromycin.
 CC The narbonolide PKS consists of a loading module, six extender modules,
 CC and two thioester domains. Four proteins make up the narbonolide PKS
 CC (PICAI, PICAI, PICAI, and PICAI). PICAI includes the loading module
 CC and extender modules 1 and 2, PICAI includes extender modules 3 and 4,
 CC PICAI includes extender module 5 and PICAI includes extender module 6
 CC and a type II thioesterase domain. The second type II thioesterase domain
 CC is found on the PICB protein. The nucleotide sequences encoding all of
 CC these proteins can be isolated in recombinant form from the recombinant
 CC cosmid pKOS023-27. Narbonolide is desosaminylated in S. venezuelae to
 CC yield narbomycin, the desosaminyl transferase enzyme is required for this
 CC conversion, and the desosamine biosynthetic genes are also found in
 CC cosmid pKOS023-27. The recombinant DNA of the invention is used to
 CC express, in transformed cells, narbonolide (or its derivatives) or other
 CC ketolides (particularly hybrids), which may then be converted (e.g. by
 CC other enzymes recombinantly expressed in the same hosts) to polyketide
 CC antibiotics or their intermediates. The antibiotics are useful in human
 CC or veterinary medicine.
 XX

SQ Sequence 3291 BP; 521 A; 1289 C; 1044 G; 436 T; 1 other;

Query Match 2.9%; Score 41.6; DB 21; Length 3291;
 Best Local Similarity 46.3%; Pred. No. 2.5;
 Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 73 ctgcagcgctagaagcaatcgctgaccccccgcgtgtccgtgagtcctcccttcgctg 132
 DB 3247 CGCTCGGCCCATTCGCGACGGCTCGATCACCGCAGCGCTCGGGCGGTCCAGGTG 3188
 QY 133 cgggaacctgctgcccgcacagcgagctgttcgagaggtggccattctcgggggcg 192
 DB 3187 CGGGCCGATCGGCAGGCTGAGGACCTCGCGCGGAGAGCTCTCGGCCCGGGAGCGAGCC 3128
 QY 193 gcggcgagggcgccccgggggtcccgggggtctcggggacttttactatctactctg 252
 DB 3127 TTCGGCGGTGCTCTCGCCCGGTAGGCGGCGAGAGGTGCACGGGTACCGGTAGTGCCT 3068
 QY 253 gccaatctgagggcgaagcagcgaggtggcgcgctacttcctcccgggcggaag 312
 DB 3067 GAGGGTGTGATCCCGGGCGGTCTGAGGTGGTGTGCCAGCTGTCTCGCGGCTCGGTGCG 3008
 QY 313 agtgcacagcagcttcttcggggggggtccagactcagcgagcagctggccaa 368
 DB 3007 CACGGTGAAGAGGTGCACACCGGGTCGGTGTGCGGCGGGTACCGGAGGCGCA 2952


```

FT FT /*tag= i
FT FT /product= "PikB gene cluster protein #9 (Y77210)"
FT FT complement (13705..15043)
FT FT /*tag= j
FT FT /product= "PikB gene cluster protein #10 (Y77211)"
FT FT 15404..15574
FT FT /*tag= k
FT FT /product= "PikB gene cluster protein #11 (Y77212)"
FT FT
FT FT WO20000620-A2.
FT FT
FT FT 06-JAN-2000.
FT FT
FT FT 25-JUN-1999; 99WO-US14398.
FT FT
FT FT 26-JUN-1998; 98US-0105537.
FT FT (MINU ) UNIV MINNESOTA.
FT FT
FT FT Sherman DH, Liu H, Xue Y, Zhao L;
FT FT
FT FT WPI: 2000-160679/14.
FT FT
FT FT P-PSDB; Y77204, Y77205, Y77206, Y77207, Y77208, Y77209,
FT FT Y77210, Y77211, Y77212, Y80998, Y80999.
FT FT
FT FT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
FT FT synthesis of methymycin and pikromycin
FT FT
FT FT Disclosure; Figure 32; 438pp; English.
FT FT
FT FT The invention relates to an isolated and purified nucleic acid segment
FT FT comprising a desosamine biosynthetic gene cluster, a fragment or its
FT FT biologically active variant, where the nucleic acid sequence is not
FT FT derived from the eryC gene cluster of Saccharopolyspora erythraea or
FT FT Streptomyces antibioticus. The invention also relates to a macrolide
FT FT biosynthetic gene cluster, or fragments thereof. The macrolide
FT FT biosynthetic gene cluster encodes proteins which synthesize methymycin,
FT FT pikromycin, neomethymycin, narbomycin or a combination of these
FT FT compounds. Recombinant or augmented cells comprising the desosamine
FT FT and/or macrolide biosynthetic gene clusters are useful for the production
FT FT of biologically active macrolides. The macrolide biosynthetic proteins
FT FT are useful for synthesis of methymycin, pikromycin, neomethymycin and
FT FT narbomycin. The alternative termination of polyketide synthesis may be
FT FT useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
FT FT monomers. The compounds produced by the recombinant host cells are useful
FT FT as biopolymers, e.g., in packaging or biomedical applications, to
FT FT engineer PHA monomer synthases or to prepare biologically active agents,
FT FT such as chemotherapeutics, immunosuppressants, agents to treat asthma,
FT FT chronic obstructive pulmonary disease as well as other diseases involving
FT FT respiratory inflammation, cholesterol-lowering agents or macrolide-based
FT FT antibiotics which are active against a variety of organisms, e.g.,
FT FT bacteria, including multi-drug resistant pneumococci and other
FT FT respiratory pathogens, as well as viral parasitic pathogens, or as crop
FT FT protection agents (e.g., fungicides or insecticides) via expression of
FT FT polyketides in plants. The present sequence represents the desosamine
FT FT biosynthetic gene cluster (pik) from Streptomyces venezuelae AICC
FT FT 15439, as given in figure 32.
FT FT
FT FT Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other;
FT FT
FT FT Query Match 2.98; Score 41.6; DB 21; Length 13613;
FT FT Best Local Similarity 46.39; Pred. No. 3.2;
FT FT Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
FT FT
FT FT Qy 73 ctgcagcgcctagagacgaatcgctgaccccccgctccgctgagtcctcccttcgctg 132
FT FT ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
FT FT Db 6845 ccgctcgcccatcctgcgcagcgctgcatcacccgagcgcctgaggcgctccaggtg 6904
FT FT
FT FT Qy 133 cgggaactgtgccccacagcagcagctgtcttcagaggtgcccatttcggcgggcg 192
FT FT ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
FT FT Db -6905 cgggccatggcaggtgaggacctgcgcggaagactctcgcccgaggagcgagcc 6964

```

```

QY 193 gccggcgaggcgcccgccgggggtcccgcgggctctggggactttctactatctactcgt 252
DB 6965 ttccggcggtgctcgcccgctagcgggcgagaggtgcacgggtacccgggtagtgct 7024
QY 253 gccaatctcgaggcgcaagagcagcgagtgcccgctgctgcctcccgggcggaagg 312
DB 7025 gagggtgctcgatgcgcggcgctcgagtgctgcgcgctcgtcgcgctcggtgctg 7084
QY 313 agtgcacagcagcagctcttcggcgggctccagactcagggcgacagctggccaa 368
DB 7085 cagcgtgaagaggtgccagaccgggtcggtgctcgggcgctgcacccgagcgcca 7140

RESULT 10
X08910
ID X08910 standard; cDNA; 1582 BP.
XX
AC X08910;
XX
DT 27-APR-1999 (first entry)
XX
DE Human FADD protein coding sequence.
XX
KW FIP; FADD interacting protein; FADD; Fas-associated protein with a
KW novel death domain; cell death; apoptosis; Alzheimer's disease;
KW Acquired Immune Deficiency Syndrome; AIDS; muscular dystrophy;
KW amyotrophic lateral sclerosis; virus; bacteria; fungus; mycoplasma;
KW protozoa; neoplasia; dysplasia; hyperplasia; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 70..596
FT /*tag= a
FT //label= Human_FADD_protein
FT
XX WC9900499-A1.
XX
XX 07-JAN-1999.
XX
XX 26-JUN-1998; 98WO-US13320.
XX
XX 03-JUN-1998; 98US-0087886.
XX 26-JUN-1997; 97US-0050792.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Chen TT, Williams LT;
XX
XX WPI; 1999-095745/08.
XX P-PSDB; W96154.
XX
XX New FADD (Fas-associated protein with a novel death
XX domain)-Interacting Protein - useful for inducing or preventing
XX apoptosis in a cell, to aid in controlling apoptosis-related
XX diseases
XX
XX Disclosure; Page 45-46; 58pp; English.
XX
XX An epitope of human FADD (Fas-associated protein with a novel
XX death domain)-Interacting protein (FIP protein) comprising amino
XX acids 348-727 of the protein described in W96153, can be used to
XX induce or prevent apoptosis in a cell. Specifically, decreasing the
XX levels of FIP348-727 prevents apoptosis. This is useful in cells
XX which are dying prematurely, eg: Alzheimer's disease, Acquired
XX Immune Deficiency Syndrome (AIDS), muscular dystrophy, amyotrophic
XX lateral sclerosis (and other muscle wasting diseases), autoimmune
XX diseases, and diseases where cells are infected with a pathogen
XX (virus, bacteria; fungus, mycoplasma or protozoa). Increasing the
XX levels of FIP 348-727 induces apoptosis which is useful in cells
XX suffering from neoplasias, dysplasias, hyperplasias, or their
XX symptoms. Purified and isolated FIP subgenomic polynucleotides are
XX useful as primers to obtain more copies of the nucleotides, and as

```

* CC probes that identify wild-type or mutant coding sequences. They are
CC also useful for expressing FIP mRNA, proteins or fusion proteins,
CC and in the generation of FIP antisense oligonucleotides and
CC ribozymes. They are also useful in expression constructs and in
CC gene delivery vehicles (optionally in combination with a condensing
CC agent) that deliver FIP mRNA or oligonucleotides, FIP proteins
CC (including variants), FIP-specific ribozymes or single-chain
CC antibodies into eukaryotic cells. This is the human FADD protein
CC gene. Human FIP protein binds to amino acids 1-110 of the human
CC FADD protein given in W96134.
XX
SQ Sequence 1582 BP; 344 A; 433 C; 483 G; 322 T; 0 other;

Query Match 2.9%; Score 41.4; DB 20; Length 1582;
Best Local Similarity 48.9%; Pred. No. 2.4; Indels 0; Gaps 0;
Matches 111; Conservative 0; Mismatches 116;
QY 21 gtggggcgacagtgggaagagccgcgccttggccggcgagtcagatgctgcagcg 80
Db 156 ggcgctgggcaagcgaagctgagcgctgcagagcgcttagacctttctccatgct 215
QY 81 cctagaagacatgctgcagcccgctgctcgtgagtcctcccttcgctgggaacct 140
Db 216 gctggagcagaacgacctggagcccgccacaccgagctcctgcgcgagctgctgcctc 275
QY 141 gctgcccgcacagcgacgctgttcgagaggtggcccttctcggcgccgcccggcg 200
Db 276 cctgcccgcacagcactgctgagaggtggcccttctcggcgccgcccggcg 335
QY 201 agcgccgcgggggtcccgcgctctctgggagcttctactcatct 247
Db 336 ggcgcgcgctgggaagaagacgtgtgacgacatttaacgtcatat 382

RESULT 11
T33937
ID T33937 standard; cDNA; 1642 BP.
XX
AC T33937;
XX
DT 22-FEB-1997 (first entry)
XX
DE FADD (Fas-associating protein with novel death domain) gene.
XX
KW Human; FADD; Fas-associating protein with novel death domain;
KW apoptosis; Fas receptor; death domain; gene therapy; antibody;
KW immunoassay; drug screening; diagnostic; AIDS; antiinflammatory;
KW antitumour; cerebroprotective; neuroprotective; ss.
XX
OS Homo sapiens.

XX Key Location/Qualifiers
FH 5'UTR 1..129
FT misc_feature 4..6
FT misc_feature 130..756
FT CDS
FT misc_feature 133..501
FT misc_feature 198
FT misc_feature 249
FT misc_feature 460..860
FT misc_feature

/note= "Region encoding death domain"
757..1642
/*tag= h
polyA_signal 1636..1641
/*tag= 1
WO9631603-A2.
10-OCT-1996.
28-FEB-1996; 96WO-US02857.
18-MAY-1995; 95US-0443982.
03-APR-1995; 95US-0416379.
(UNMI) UNIV MICHIGAN.
Dixit VM, O'Rourke K;
WP1; 1996-465026/46.
P-PSDB; W03653.
FADD protein that binds to cytoplasmic region of Fas receptor - for
identifying inhibitors of Fas-associated apoptosis useful for
treating e.g. AIDS, leukaemia, stroke, etc
Example 1; Fig 2A-B; 96pp; English.
The sequence encodes FADD (Fas-associating protein with novel death
domain), which binds the cytoplasmic region of a Fas receptor, and
modulates apoptosis induced by activation of the receptor by ligand
binding. The cDNA has been isolated using a yeast two-hybrid system
which screens for proteins interacting with the Fas cytoplasmic
domain. A GAL4 DNA-binding domain has been fused to the human Fas
antigen cytoplasmic tail to form a bait plasmid, which is used with
a prey plasmid, containing a human B-lymphocyte cDNA library fused to
the GAL4-activation domain, to co-transform yeast cells.
Overlapping clones 8 and 15 have been isolated, and the full-length
cDNA has been isolated from a human umbilical vein endothelial cell
(HUVCE) library using clone 15 as a probe. The cDNA has an
in-frame stop codon 130 bp upstream of the initiator Met. The
encoded protein contains a death domain, with interactions with the
death domain of Fas. The DNA may be used in gene therapy, and the
protein or a corresponding antibody may be used to screen for
agents modulating FADD pathway cellular functions and Fas-associated
apoptosis, for use in therapy of e.g. AIDS, inflammation, etc.
leukaemia, myocardial infarction, degenerative disease, etc.
SQ Sequence 1642 BP; 354 A; 448 C; 508 G; 332 T; 0 other;

Query Match 2.9%; Score 41.4; DB 17; Length 1642;
Best Local Similarity 48.9%; Pred. No. 2.4; Indels 0; Gaps 0;
Matches 111; Conservative 0; Mismatches 116;
QY 21 gtggggcgacagtgggaagagccgcgccttggccggcgagtcagatgctgcagcg 80
Db 216 ggcgctgggcaagcgaagctggagcgctgacagcgccctagaccttctccatgct 275
QY 81 cctagaagacatgctgcagcccgctgctcgtgagtcctcccttcgctggggacct 140
Db 276 gctggagcagaacgacctggagcccgccacaccgagctcctgcgcgagctgctgcctc 335
QY 141 gctgcccgcacagcgacgctgttcgagaggtggcccttctcggcgccgcccggcg 200
Db 336 cctgcccgcacagcactgctgagcgctgagaggtggcccttctcggcgccgcccggcg 395
QY 201 agcgccgcgggggtcccgcgctctctgggagcttctactcatct 247
Db 396 ggcgcgcgctggggaagaagacctgtgtgcagcatttaacgtcatat 442
RESULT 12

Z39042/c
 ID Z39042 standard; cDNA; 1170 BP.
 AC Z39042;
 XX
 XX 25-FEB-2000 (first entry)
 DT
 DE Human fishboy G-protein coupled receptor encoding cDNA.
 XX
 KW Human; fishboy; G-protein coupled receptor; therapy; diagnosis;
 KW signal transduction; HIV; infection; pain; cancer; diabetes; obesity;
 KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;
 KW hypotension; hypertension; urinary retention; stroke; osteoporosis;
 KW angina pectoris; myocardial infarction; migraine; ulcer; allergy;
 KW benign prostatic hypertrophy; vomiting; dyskinesia; psychotic disorder;
 KW neurological disorder; Huntington's disease; Tourette's syndrome; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1170
 FT /*tag= a
 FT /product= "Fishboy"
 FT /note= "a G-protein coupled receptor"
 XX
 XX W09942484-A1.
 XX
 XX 26-AUG-1999.
 XX
 XX 16-FEB-1999; 99WO-US03246.
 XX
 XX 20-FEB-1998; 98US-0075626.
 XX 13-AUG-1998; 98US-0133132.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 XX Elshourbagy N, Lane PA, Tsui P, Vawter L;
 XX
 XX WPI; 1999-527456/44.
 XX P-PSDB; Y57456.
 XX
 XX A new G-protein coupled receptor, Fishboy used to treat, e.g. benign
 XX prostatic hypertrophy -
 XX
 XX Claim 2; Page 34-35; 49pp; English.
 XX
 XX The present sequence encodes a human G-protein coupled receptor,
 XX designated Fishboy. Fishboy products can be used to treat subjects in
 XX need of modulated Fishboy expression or activity. Diseases or disorders
 XX to be treated include bacterial, protozoal, fungal and viral,
 XX particularly HIV-1 or HIV-2 infections, pain, cancers, diabetes, obesity,
 XX anorexia, bulimia, asthma, Parkinson's disease, acute heart failure,
 XX hypotension, hypertension, urinary retention, osteoporosis, angina
 XX pectoris, myocardial infarction, stroke, ulcers, allergies, benign
 XX prostatic hypertrophy, migraine, vomiting, psychotic and neurological
 XX disorders and dyskinesias such as Huntington's disease or Tourette's
 XX syndrome.
 XX
 XX Sequence 1170 BP; 137 A; 409 C; 415 G; 209 T; 0 other;
 SQ

Query Match 2.3%; Score 41.2; DB 20; Length 1170;
 Best Local Similarity 44.3%; Pred. No. 2.5;
 Matches 210; Conservative 0; Mismatches 263; Indels 1; Gaps 1;
 QY 8 tggcggtggccgctggggcgacagtgggaagagccgcgcctgggcccggcgagta 67
 Db 815 TGGTAGGGGGCCGAGCAAGCCGAGGACGACGATGGCGCTACACAGCGGCCACC 756
 QY 68 ggatgctgcagcgcctagaagagcaatgcgtcagcccggtgctgcgtgagtcctcc 127
 Db 755 CGCGCCCGTGGCGCGGAGCCCGAGCGCGGCCCGCGAGCGCTGCCAGCGTACGCTG 696

QY 128 cgctggggaactgctgccccgcacagcgacagctgcttcgagagtggtgcccattctc-gg 186
 Db 695 TAGCAGCCGAGCATCAGCCCGAAGAGCAAGAGCGGTGAGAGTCTCAGGGCTCAGG 636
 QY 187 cggggcgccggcgagggcgccccgggggtcccgggcgctctctggggactttctactc 246
 Db 635 TGGGGCGGGCGCTGGACGCGGACGGGTGGCACAGCTGGCATACGCGTCCCTCCACAGG 576
 QY 247 tacctggccaatctggagggccaagagcgaggtggcgcgctgctcctccccggggc 306
 Db 575 TGGCGGTAGACGCGCGCGCGGACGCGAGCAACAGGGGGCCAGCAGACCGCCAGCAGC 516
 QY 307 cgaagagtgcccaacgacgagctcttcggggcggtccagactcagcgacagtggtg 366
 Db 515 AGCGCGCGGGCAGGGCGGGCTGGCAGCGCGGAGCGCCAGGAGGGGGGTGACTGCG 456
 QY 367 asgtggccatcttcagcccatgcaacgacagtgacgacgtgacgacacttctccccggggca 426
 Db 455 AGGCAGCGCTGACGGCTGAGCAGCGCGGTGAGCAGCAGCGTGGCGTACATGCTGAGCGG 396
 QY 427 agtactgtgacacatgtaccagctcaccaggtcccaaggccccccgcccacacctctgg 480
 Db 395 CACAGTAGTACACCGCTTGACGCGCGCTGGCGCCGCTGGCCAGCGCGCTGCCGG 342
 RESULT 13
 X53491/c
 ID X53491 standard; DNA; 114955 BP.
 XX
 AC X53491;
 XX
 DT 05-JUL-1999 (first entry)
 XX
 DE Human adenosine A1 receptor antisense oligonucleotide fragment.
 XX
 KW Antisense oligonucleotide; multiple target; antisense treatment;
 KW impaired respiration; inflammation; lung disease;
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
 KW acute asthma; allergy; asthma; impeded respiration;
 KW respiratory distress syndrome; pain; cystic fibrosis;
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
 KW prostate cancer; ss.
 XX
 OS Synthetic.
 OS
 XX W09913886-A1.
 XX
 PD 25-MAR-1999.
 XX
 PF 17-SEP-1998; 98WO-US19419.
 XX
 PR 09-JUN-1998; 98US-0093972.
 PR 17-SEP-1997; 97US-0059160.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 XX Nyce JW;
 PI
 XX WPI; 1999-229400/19.
 DR
 XX New antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction
 XX
 PS Disclosure; Page 37; 120pp; English.
 XX
 CC The specification describes antisense oligonucleotides (X52869-X55271)
 CC directed against at least 2 mRNAs selected from target genes, coding and
 CC non-coding regions of RNAs corresponding to target genes, gene
 CC initiation codons, genomic flanking regions, intron-exon borders, the
 CC 5'-end, the 3'-end and the juxta-section between coding and non-coding

KW MORT-1; HFL1; FAS/AP01 receptor; FAS-R; tumour; cancer; HIV;
 KW mediator of receptor toxicity; gene therapy; ss.
 XX Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 1..771
 FT /*tag= a

PN W09618641-A1.

XX 20-JUN-1996.

XX 14-DEC-1995; 95WO-US16542.

XX 16-JUL-1995; 95IL-0114615.

PR 15-DEC-1994; 94IL-0112022.

PR 19-FEB-1995; 95IL-0112892.

XX (WEIN/) WEINURZEL H.

PA (YEDA) YEDA RES & DEV CO LTD.

XX Boldin M, Mett I, Varfolomeev E, Wallach D;

XX WPI: 1996-300569/30.

DR P-PSDB; R98346.

XX MORT-1 protein capable of interacting with FAS-R intracellular
 PT domain - useful for modulating FAS-R ligand effect on cells and
 PT treating, e.g. tumour cells and HIV-infected cells

PS Claim 3; Fig 4; 72pp; English.

XX A cDNA clone (T30372) codes for MORT-1 (R98346) (Mediator of
 CC Receptor Toxicity), also designated HFL1, a novel protein that binds
 CC to the intracellular domain (Fas-IC) of the Fas ligand receptor FAS-R
 CC (or FAS/AP01), and is capable of modulating the function of Fas-R.
 CC It was obtd. from HeLa cells using a yeast 2-hybrid screen and 2-hybrid
 CC beta-galactosidase expression system. The cDNA can be used for produ.
 CC of recombinant MORT-1 using transformed host cells. It can also be
 CC used to modulate the FAS-R ligand on cells carrying an FAS-R and to
 CC develop methods for the gene therapy of e.g. cancer and HIV infection.

XX Sequence 1701 BP; 382 A; 459 C; 517 G; 343 T; 0 other;

Query Match 2.8%; Score 40; DB 17; Length 1701;

Best Local Similarity 49.1%; Pred. No. 4.9;

Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 32 agtgggagagagcccgccctggcgccgagtcagtgatgcagcgccctagagagc 91

Db 242 agcgcaagctggagcgctgcagagcgccctagacattctctccatgctgctggagcaga 301

QY 92 aatgctgcagcccccgggtctccgtgagtcctccctctgctgggagacctgctgccccgca 151

Db 302 acgacctggagcccgggcacaccagatcctctgcgagctgctgcctccctgcggcgcc 361

QY 152 cagcgagctgctcagagagtggtgccattctcgccggcgccggcgagggcgcccg 211

Db 362 acgacctgctgcggcgctgcagacgttcgagcgggcgccggcgccggcgcccg 421

QY 212 ggggtcccgccggcgctctggggactttctactatct 247

Db 422 gggagaagacctgtgtgcagcatttaagtcatat 457

Search completed: December 7, 2000, 18:10:10
 Job time: 26043 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 13:43:21 ; Search time 3361.48 Seconds
(without alignments)
1847.608 Million cell updates/sec

Title: US-09-434-708-3

Perfect score: 1422

Sequence: 1 atggctctgcggtggccccc.....ccaggaccctgcccgcc 1422

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba1.*

2: gb_ba2.*

3: gb_om.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pl1.*

8: gb_pl2.*

9: gb_pri.*

10: gb_pr2.*

11: gb_pr3.*

12: gb_ro.*

13: gb_sy.*

14: gb_un.*

15: em_fun.*

16: em_hum1.*

17: em_hum2.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vi.*

30: gb_htg1.*

31: gb_htg2.*

32: gb_in1.*

33: gb_in2.*

34: em_ba1.*

35: em_ba2.*

36: em_hum3.*

37: em_hum4.*

38: gb_pr4.*

39: gb_htg3.*

40: gb_htg4.*

41: gb_htg5.*

42: gb_htg6.*

43: gb_htg7.*

44: em_htg1.*
45: em_htg2.*
46: em_htg3.*
47: em_hum5.*
48: gb_pl3.*
49: gb_pr5.*
50: gb_htg8.*
51: gb_htg9.*
52: gb_htg10.*
53: gb_htg11.*
54: gb_htg12.*
55: gb_htg13.*
56: gb_htg14.*
57: gb_in3.*
58: gb_htg15.*
59: gb_htg16.*
60: gb_htg17.*
61: em_htg4.*
62: em_htg5.*
63: em_htg6.*
64: em_htg7.*
65: em_hum6.*
66: gb_htg18.*
67: gb_htg19.*
68: gb_htg20.*
69: gb_htg21.*
70: gb_htg22.*
71: gb_htg23.*
72: gb_vil.*
73: gb_vil2.*
74: gb_ba3.*
75: em_htg8.*
76: em_htg9.*
77: em_htg10.*
78: em_htg11.*
79: em_htg12.*
80: em_htg13.*
81: em_htg14.*
82: em_htg15.*
83: em_htg16.*
84: em_htg17.*
85: em_htg18.*
86: em_htg19.*
87: em_htg20.*
88: em_htg21.*
89: em_htg22.*
90: em_htg23.*
91: gb_pr6.*
92: gb_pr7.*
93: gb_sts1.*
94: gb_sts2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1418.8	99.8	1584	11	AF117646	AF117646 Homo sapi
2	1415.6	99.5	1575	9	AB028645	AB028645 Homo sapi
3	1132.8	79.7	1446	11	AF117647	AF117647 Homo sapi
4	340	23.9	3090	91	HSCCBL	X57110 Human mRNA
5	329.2	23.2	2661	57	DMU87925	U87925 Drosophila
6	327.6	23.0	2529	57	DMA3175	AJ233175 Drosophila
7	323.6	22.8	3241	91	HSU26711	U26711 Human cbl-b
8	323.6	22.8	3354	91	HSU26712	U26712 Human cbl-b
9	323.6	22.8	3982	91	HSU26710	X57111 Mouse mRNA
10	270.6	19.0	2808	12	MMCCBL	J04169 Mouse Cas N
11	248.2	17.5	2711	72	CASNS1	AC073760 Mus muscu
12	214.2	15.1	237653	60	AC073760	

```

13 199.4 14.0 2230 33 CEDNASL11
C 14 179.2 12.6 36890 41 AC014994
15 179.2 12.6 139226 39 AC010033
C 16 179.2 12.6 281936 32 AE003555
C 17 139.8 11.2 208239 51 AC021988
18 127.2 8.9 1758 4 AF237765
C 19 98 6.9 32559 33 CELM02A10
C 20 81.2 5.7 173038 70 AP001003
C 21 81.2 5.7 177595 70 AP001557
C 22 76.4 5.4 169523 41 AC015788
C 23 76.4 5.4 179680 11 AC016138
24 56.8 4.0 4855 2 SA003114
25 51.8 3.6 11212 53 AC025047
26 49.4 3.5 17183 1 AE001986
27 49 3.4 63734 1 AF127374
C 28 48.6 3.4 87283 51 AC022856
29 47.8 3.4 40352 2 SCC61A
30 47.6 3.3 12801 2 SCF6
31 47.2 3.3 107444 66 AC078780
C 32 46.6 3.3 2510 73 SH1UL2XA
33 46.6 3.3 3797 73 PVULSGENE
34 46.6 3.3 9382 73 SH1ULGNS
35 46.4 3.3 16951 2 MSU46844
C 36 46.4 3.3 208918 60 AC074335
37 46.2 3.2 807 91 HSKCNQF01
38 45.6 3.2 34962 2 SCG20A
C 39 45.2 3.2 149402 67 AL157400
C 40 45 3.2 7913 3 BOVCMPR
C 41 44.8 3.2 15559 1 AF074603
C 42 44.8 3.2 40356 2 SCL6
C 43 44.8 3.2 169054 52 AC023767
C 44 44.6 3.1 9999 1 AE001942
C 45 44.2 3.1 10419 1 AE001865

ALIGNMENTS

RESULT 1
AF117646 1584 bp mRNA PRI
LOCUS Homo sapiens long CBL-3 protein (CBL-3) mRNA, alternatively spliced
DEFINITION isoform, complete cds.
ACCESSION AF117646
VERSION AF117646.1 GI:4959420
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1584)
AUTHORS Keane,M.M., Ettenberg,S.A., Nau,M.M., Banerjee,P., Cuello,M.,
Penninger,J. and Lipkowitz,S.
TITLE cbl-3: a new mammalian cbl family protein
JOURNAL Oncogene 18 (22), 3365-3375 (1999)
MEDLINE 99289203
REFERENCE 2 (bases 1 to 1584)
AUTHORS Keane,M.M., Ettenberg,S.A., Nau,M.M., Banerjee,P., Cuello,M.,
Penninger,J. and Lipkowitz,S.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-1999) Genetics Dept, Medicine Branch, DCS, NCI,
Bethesda Naval Hospital, Bethesda, MD 20889, USA
FEATURES
source
1..1584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.2"
/cell_line="CF-PAC1"
/tissue_type="pancreatic adenocarcinoma"
1..1584
/gene="CBL-3"
64..1488
CDS

```

```

/feature="CBL-3"
/feature="long form, alternatively spliced; cbl family
protein; structurally related to the sli-1 protein,
encoded by the cbl gene of Caenorhabditis elegans"
/codon_start=1
/product="long CBL-3 protein"
/protein_id="RAD34341.1"
/db_xref="GI:4959421"
/translation="MALAYPWQWEERARALGRAVRLQRLEEQVDPRLSVSPSL
RDLIPRAQLUREVAHRRAGGPGGPGGDFLLIYLANLEAKSRQVAALLPRG
RRSANDLFRAGSLRLKALAIIFSHMAELHALFPYGYCHMYQLKAPAFIFW
RESCGACVLTPWAEFSLGTCHPEVPGCTALARTIDTICSHYVIFEDVTRLF
QPWTLKLNWOLLAVNHPGYMAFTYDEVQERLQACRDKPGSVIFRSPRLGWAIG
YVSSDGSILQTPANKPLSOVLLEGOKDGLYDPDKTHNPDLTELGOAEPQRIHVS
BEOLQLYAMDSTFELCKIAESKNKYIEPCGHLHLLCSCLAAWQSDSDTCCFRCNE
IKGWEAVSITQFYCATAEQSDSSDOEGRELELQGVPLSAPPLPRPPLPRKPRNA
QPKVRLKNGSPPAALGPDPA"
BASE COUNT 303 a 538 c 482 g 261 t
ORIGIN
Query Match 99.8%; Score 1418.8; DB 11; Length 1584;
Best Local Similarity 99.9%; Pred. No. 5.1e-242;
Matches 1420; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 atggctctggcgggtggccccgtggggcgacagtgggagagggccccggcctggcgcg 60
DB 64 ATGGCTCTGGCGGTGGCCCCGTGGGGCGACAGTGGGAGAGAGCGCGCCTGGCGCG 123
QY 61 gaagtcaagagctgcagcgcctagaagagcaatcgctgaccccccgctgtccgtgagt 120
DB 124 CCAATCAGCATGCTGCACGCGCTAGAGAGCAATGCGTCCGACCCCGCGCTGCTCGT 183
QY 121 cccctctgcctggggacctctctcccgacagcgagctgttcagaggtggcccat 180
DB 184 CCCCCTTGGCTGGGGACCTGCTGCCCGGCACACGGCAGCTGCTTCAGAGGTGGCCAT 243
QY 181 tctggcgggcgccggcggaggcgccccgggggtcccgggcgctctggggactttcta 240
DB 244 TCTGGCGGGCGGGCGGGAGGCGGGCGGCGGCGGCTCGGGGCTGCTGGGACTTTCTA 303
QY 241 ctcatctacctggcdaatctgagagccaaagagcagatggcgcgtgcctccccc 300
DB 304 CTCATCTACCTGGCCATCTCTGGAGGCAAGAGAGGAGGAGTGGCGCGCTGCTGCTCC 363
QY 301 cggggcggaagagtgccaacgacgagctctccggcgggggtccacgactcagcgacag 360
DB 364 CGGGCGCAAGAGTGGCCAAACGACGAGCTCTTCGGGGCGGGTCCAGACTCAGCGACAG 423
QY 361 ctggcgaagctggcgcacatcttcagccacatgcacgacagctgcagcactctcccc 420
DB 424 CTGGCAAGCTGGCCATCATCTTCAGCCCATGCACGACGAGTGCACGACTCTTCCCC 483
QY 421 ggggcaaaagtactgtgacacatgtaccagctcaccagcctcccgccccccacacctctgg 480
DB 484 GGGGGAAGTACTGTGGACACATGTACCAGCTCACCAGGCGCCCGCCACACCTTCTGG 543
QY 481 agggaaagtctcgagagcccggtgtgtctccctggcgtgagttgagttccctccctgggc 540
DB 544 AGGAAAAGTTTCGGAGCGCGGTGTGTCTGCCCTGGCTGAGTTTGAAGTCCCTCCCTGGGC 603
QY 541 acctggcacacctgtggaaacagcgtgcacagcctggccttgcgacacacattgacctc 600
DB 604 ACCTGCCACCTGTGGAACACGAGCTGCACAGCCTGGCCTTGGCCACCATTTGACCTC 663
QY 601 acctgacggcgacgtgtgctcatttcttgcagtttcagtttcacccaggtcttttcagcgc 660
DB 664 ACCTGCAGCGGCGACGCTGCTCATCTTCAGTTCGACGCTCTTCACCGAGGCTCTTTACGCCA 723
QY 661 tggccaacacctcctcaagaactggcagctcctctgagttcagtttcacccaggtctacatggcc 720
DB 724 TGGCCACACCTCTCAGAACTGGCAGCTCCTGGCAGTCAACCCAGCCAGCTACATGGCC 783

```

```

Qy 721 ttcctccatgatgaggtccaaagagctgtgaaggctgcaggagcaggaagccaggcagt 780
Db 784 TCCCTCACCATATGATGAGTCCAAAGAGCGTCTCGAGGCTGCAGGAGCAAGCCAGGCAGT 843
Qy 781 tacatcttcggccagctgtactgcgtctggggcagtggtggccatcggtctatgtgagctca 840
Db 844 TACATCTTCGGGCCAGCTGTACTCGCTGGGCGAGTGGGCCATCGGTATGTGAGCTCA 903
Qy 841 gatggcagatctgcagaccatctccctgcgaacaaacccctctccagggtgctcctggag 900
Db 904 GATGGCAGATCTCGCAGACCATCCCTGCGCAACAAACCCCTGCCAGGTGCTCCTGGAG 963
Qy 901 gacagaagagcggtcttcacctctaccagatggaagaccccaacccacacacctgact 960
Db 964 GGCACAAGGACGGCTTCTACCTCTACCCAGATGGAAGACCCACACCCAGACCTGACT 1023
Qy 961 gagctcggcaggcagaacccccagcagcgcacatccacgtctcagagagcagctgcagctc 1020
Db 1024 GAGCTGGCAGCAGCAAGACCCAGCAGCGCATCCAGCTGTCTAGAGGAGCAGCTGCAGCTC 1083
Qy 1021 tactggccatggactccacattttagctctgcaagatctgtctgagagcacaagat 1080
Db 1084 TACTGGGCCATGACTCCACATTTGACCTCTGCAAGATCTGTCTGAGAGCAGCAGGAT 1143
Qy 1081 gtgaagattgagcgtgcgggcaacctgtctgcagctgctgctggtcctcggcagcac 1140
Db 1144 GTCAAGATTGAGCGCTGCGGCGACCTGCTGTGAGCTGTGCTGCTGCTGCTGCGCAGCAC 1203
Qy 1141 tcggaacagcagacctgcctctctccgtcgagatcaaggctggagagcgtgact 1200
Db 1204 TCGGACAGCAGCAGCTGCCCCCTTCTCCGCTGCGAGATCAAGGCGCTGGAGGCGGTGAGT 1263
Qy 1201 attaccagttccacggtcaggtactgtctgagagctcagggaacagcagtgaccaggaa 1260
Db 1264 ATCTACAGTTCTACGGTCAGGCTACTGTCTGAGGATCAGGGAACAGCAGTACCCAGGAA 1323
Qy 1261 ggacggagttgagctggggcagggtgccccttctgctcctcattgccccacagcaca 1320
Db 1324 GCGAGGAGTTGAGCTGGGCGAGTGGCCCTTTTGGCTCTCTCATTTGCCCTCCACGGCCA 1383
Qy 1321 gactccccccaggagccagaaatgccagccagaaagttagactctaaaggaggaac 1380
Db 1384 GATCTGCCCCCAGGAGGCCAGAAATGCCAGCCGAAAGTAGACTCTTAAGGGGGAAC 1443
Qy 1381 tccctccagctgctgggacccccaggacctgcctcctccctggcc 1422
Db 1444 TCCCTCCAGCTGGCTGGGACCCAGACCCAGACCTTGGCCCGGCC 1485

RESULT 2
AB028645 1575 bp mRNA PRI 10-NOV-1999
LOCUS Homo sapiens mRNA for Cbl-c, complete cds.
DEFINITION AB028645
ACCESSION AB028645
VERSION AB028645.1 GI:6291531
KEYWORDS Cbl-c.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kim,M., Tezuka,T., Suzuki,Y., Sugano,S., Hirai,M. and Yamamoto,T.
TITLE Molecular cloning and characterization of a novel cbl-family gene,
Cbl-c
JOURNAL Gene 239, 145-154 (1999)
REFERENCE 2 (bases 1 to 1575)
AUTHORS Kim,M., Tezuka,T. and Yamamoto,T.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-1999) to the DDBJ/EMBL/GenBank databases, Tadashi
Yamamoto, Institute of Medical Science, University of Tokyo,
Department of Oncology, Shirokanedai 4-6-1, Minato-ku, Tokyo
108-8639, Japan (E-mail:kyamamoto@ims.u-tokyo.ac.jp,
Tel:++81-3-5449-5301, Fax:++81-3-5449-5413)

```

```

FEATURES
    source          Location/Qualifiers
     1..1575
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /map="19q13.2-3"
     /chromosome="19"
     42..1466
     /gene="cbl-c"
     /gene="cbl-c"
     42..1466
     /gene="cbl-c"
     /codon_start=1
     /product="Cbl-c"
     /protein_id="BAA86298.1"
     /db_xref="GI:6291532"
     /translation="MALAVAPGWQWEERALGRVRLQRLLEEQCVDRLSYSPFSL
RDLLPRTAQLLREVAHSRAAGGGPGGSGDFLLIIVLANEAKSRQVALLPRRG
RRSANDLFRAGSELRLQAKLAIIFSHMAELHALFPGKYCGHMYQLTKAPAHFW
RESCGARCLVPNREFESLGTCHPVEPGCTALALRTIILDTSGSHVSIREFVFTLE
QPMPTLLKNWLLAVTHPGTMFLTYDEQVRLQACDRKPGSIFRPSCTIRUGQWALG
YVSDGSILOTIPANKPLSOVLLEKQDGFYIPYDKGTHNPDLTELGQAPQQRHVS
EEQLQYWMADSTFEICKIAESNKDKIEPCGHLKSCCLAAQWSDSQTPFCRCE
'IKGWEAVSYOFYQATAEDSGNSDQEGRELELGVPLSAPLPPRPDLPPRKPRNA
QPKVRLKGNSPPALGPDPA"
BASE COUNT      316 a , 527 c , 474 g , 258 t
ORIGIN
Query Match      99.5%; Score 1415.6; DB 9; Length 1575;
Best Local Similarity 99.7%; Pred. No. 1.9e-241;
Matches 1418; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 atggctctggcggggtggccctctggggcgacagtgggaagagccgcgcctcggcgcg 60
Db 42 ATGGCTCTGGCGGGTGGCCCTGGGGCGACAGTGGGAAGAGCCCGCCGCTGGCGCGG 101
Qy 61 cagctcagatgctgcagcgcctagaagagcaatgcgcaccccccgcgtccgtgagt 120
Db 102 GAGTCAAGATGCTGCAGCGCTAGAACAGCAATGCGTCGACCCCGGCTGTCGTGAGT 161
Qy 121 cccctctcgtcggggacctgctccccgcacagcagctgcttcagagaggtggcccat 180
Db 162 CCCCCTTCGCTGGCGGACCTGTCGCCCGCACAGCGAGCTGCTTCGAGAGGTGGCCAT 221
Qy 181 tctggcggggcgccggcgaggcgcccgggggtcccgcgactctgggagatttcta 240
Db 222 TCTCGCGGGGCGCGCGCGGCGGCGGCGGCGGCTCCCGCGGCTCTGGGGAGCTTCTA 281
Qy 241 ctcatctacctggccaatctgtgaggccaaagacagcaggcaggtggccgctgctccc 300
Db 282 CTCATCTACCTGGCCAAATCTGGAGGCCCAAGAGCAGGACAGTGGCGGCTGCTGCTCCC 341
Qy 301 gggggccaaaggagtgcacaagcagagctcttcggggcgaggctccagactcaggcgacag 360
Db 342 CGGGGCCGAAAGGAGTGCCACGACGAGCTCTTCGGGGCGGCTCCAGACTCAGCGGACAG 401
Qy 361 ctggccaaagctggcccatctattcagccacacatgcacgcagagctgcagcactcttccc 420
Db 402 CTGGCCAAAGCTGGCCCATCATCTTCAGCCACATGACGACGAGCTGCAGCGACTCTTCCC 461
Qy 421 gggggcaaaagtactgtggacacatgtaccagctcaacaaaggccccccgcacaccttgg 480
Db 462 GGGGAAAAGTACTGTGGACATGTACCAAGTGTACCAAGGCCCCCGCCACACCTTCTGG 521
Qy 481 agggaaagtgcggagcccggtgtgtgctgcctgggctgagtttggctcctctgggc 540
Db 522 AGGGAAGTGTGGAGCCCGGTTGTGTGCTGCTGGCTGAGTTTGGTTGAGTCTCTCTGGGC 581
Qy 541 acctgcacacctgtggaaccaggctgcacagccctggccttgcgccaccacctgacctc 600
Db 582 ACCTGCCACCTGTGGAAACAGGCTGCACAGCCCTGGCTTGGCGACCAACCATTCACCTC 641
Qy 601 acctgcagcggcagctgtccatcttcagtttcagcttcaccagcgtctcttaccagca 660

```



```

Db 650 ATCTTCCAAAGTGGACTCTTTCAGGAGACACATTCGGATTACTAAAGCAGATGCTGGC 709
QY 472 acctctgagggaaagtgtggagcccggtgtgtgtgctgcccctggctgagttgagttcc 531
Db 710 GAAATTTGGAGAAAGCTTTTGGGAAAGACATAGTCCCTTGAAGAGATTTCGACAG 769
QY 532 ctctgggacactgccaccctgtggaaccagctgcacagccctggccttgcgaccacc 591
Db 770 GCTCTACATGAAGTGCATCCCATCAGTCTCTGGCTGGAGGCCATGGCTCTGAAATCCACT 829
QY 592 attgacctcaactgcagggcgcaagtgtccattcttgagttgcagctcttaccagcttc 651
Db 830 ATTGATCTGACCTGCAATGATTATATTTCGGTTTGTGAATTTGACATCTTTACCCGACTC 889
QY 652 tttagcctatggcccaactcctcctcaagactggcagctcctggcagtcacaccaccaggc 711
Db 890 TTTAGCCCTTGCTCTCTTCTCAGGAATTGGAACACCTTGTGTAATCATCTCTGGC 949
QY 712 tacatggccttctcactacatgataggtccaaagagcgtcgcaggccctgcaggagcaag 771
Db 950 TACATGGGTTTTTGACGATGACGAAGTGAAGCTCGGCTCCAGAAATTCATTCAAAA 1009
QY 772 ccaggcagttacattctccgcccagctgtactgcctggggcagtgggccatcggtat 831
Db 1010 CCTGCACTTATATCTCCGGCTGAGCTGTACTGCTGTGGGTCACTGGGCTATTGGGTAT 1069
QY 832 gtgagctcagatggcagatcctgcagaccatccttgcacaaacccctgtcccaggtg 891
Db 1070 GTTACTGCTGATGGGAACATCTCCAGACAATCTCCATAAACCCTCTCTCCAAAGCA 1129
QY 892 ctctcggaggacagaagcagcttctacctctaccatgatgaagaccacaccca 951
Db 1130 CTGATTGATGCTTCAGGAAGGTTCTATTGTTTCTGATGACCAAAATCAGAACTCT 1189
QY 952 gacctgactgagctcggcagcagaaaccccgagcagcgcacatccagtgctcagaggagcag 1011
Db 1190 GATCTGACTGGCTTATGTGAACCAACTCCCAAGACCATTATCAAAAGTGACCCAGAACAA 1249
QY 1012 ctgagcttacttggccatgactacacattgagcttgcagatctgtctgagagc 1071
Db 1250 TATGAATTTACTGTGAGATGGGCTCCACATTCACATATGTAATAATATGTGTGTAATA 1309
QY 1072 aacaagatgtgaagattgagcgtgcgggcaccctgctgcagctgctcctggtgctg 1131
Db 1310 GATGAAGGATGTAAGATTGAGCCCTGTGGACACCTCATGTGCATCTCTCTATATCC 1369
QY 1132 tggcagcactggacagccagacctgccccttctgcccctgcagatcaaggctggag 1191
Db 1370 TGGCAGGAATCAGAAATCAGAGTCAAGGCTGCTCTTTCGGGATGTGAAATTAAGGTACTGAA 1429
QY 1192 gccgtgagttactaccagtt 1211
Db 1430 CCATCTGTGTAGTCCGTT 1449

RESULT 5
DMU87925
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 2661)
Meisner,H., Daga,A., Buxton,J., Fernandez,B., Chawla,A.,
Banerjee,U. and Czech,M.P.
Interactions of Drosophila Cbl with epidermal growth factor
receptors and role of Cbl in R7 photoreceptor cell development

```

```

JOURNAL Mol. Cell. Biol. 17 (4), 2217-2225 (1997)
MEDLINE 97220014
REFERENCE 2 (bases 1 to 2661)
AUTHORS Daga,A., Meisner,H., Banerjee,U. and Czech,M.P.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1997) Molecular, Cell and Developmental Biology,
University of California, Los Angeles, CA 90095, USA
FEATURES
Location/Qualifiers
1..2661
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
314..1660
/gene="D-cbl"
314..1660
/gene="D-cbl"
/codon_start=1
/protein_id="AAC47487.1"
/db_xref="GI:1842453.1"
translation="MATGSGTRVOSQPKIFPSLSKLHGAISEACVSQRLSTDKRTL
EKTWKLMDVVYKLCQOPKMNKNSPFIIDLPDYQRLRLIYKSKEDOMELLHANH
FNVTINLMKRCRAIKFKEGKEKMFEDNSHYRNLTSLVFSHMLSKALTFPG
VFAGDPRIITRADAADFKNFNGNSTLVPWKIFQELSKVPIITSGUEALAKTIDL
TCNDPISNFEFDFTRLPQWTLRNQIILAVTHPGYVAFILVDEYKARLQYLKA
GSYVRLSCIRLQWATGYVTAEGEILQITPQKSLCOALLDHREGFYLYPQQAIN
PDLSSAVOSPTEHDITVBOEYELCEGMSFTQCKICAENDKDIRIEPCGHLLCTPC
LTSQVDSBEGQCPFCRAEIKGTQIIVDAFDPKQHNRYNTNGRQQQBEDDTVE"
BASE COUNT 704 t 681 c 582 g 690 t 4 others
ORIGIN
Query Match 23.24; Score 329.2; DB 57; Length 2661;
Best Local Similarity 61.7%; Pred. No. 3.7e-49;
Matches 542; Conservative 0; Mismatches 333; Indels 3; Gaps 1;
QY 316 gcaacagcagtccttccggggctccagactccagcagcagctggcccaagctggcc 375
Db 683 GCAAGGAGAGATGTTTCGAGGAGAACTCCACATACCGCCGCAATCTCACCAAGCTCAGC 742
QY 376 atcatcttcagccacatgcacgcagagctgcagcactcttcccccggggcaagtaactgt 435
Db 743 CTGGTCTTCTCCACATGCTCAGCGAACTGAAGGCCATATTCGCCAAGGCTGTCTTTGCC 802
QY 436 ggacacatgtaccagctcaccaggcccccgcacacactcttgaggagaaagtggga 495
Db 803 GGGGATCAATTCGGATCACCAGGCGATGGGCTGACTTTTGAAGAGCAACTTCGGT 862
QY 496 gcccggtgtgtgctgcccctgggtgtgagtttgagtccttccctggcaccctgctg 555
Db 863 AACAGCAGATTGGTTCCTGGAAATCTTCGGCAGGAGCTTAGCAAGATACATCCCAT 922
QY 556 gaaccagcctcacagccctgcttgcgaccaccattaccctaccctcagcagggcac 615
Db 923 ATCTCCGGCTTGGAGGCCATGGCCCTAAAGACCACTATCGATCTTACTTGAACGACTTC 982
QY 616 gtgtcatcttctgagttgcagctcttaccaggtcttccagcagctggcaccactcctc 675
Db 983 ATTTCCAACTTCGAGTTGAGCTTCTACAGCCCTCTTCCAGCCCTTCTCCAGCTGGTGACACTGCTA 1042
QY 676 aagaactggesctctcctggagtcacaccaccaggtacatggccttctcaccatgat 735
Db 1043 CGCAACTGGCAGATTCTGGCCGTACACATCCGGGCTACGTGGGGTCTTCACATAGCAG 1102
QY 736 gaggtccaagagcgtctgcagccctgcaggagcaagcagcagcagcttaccatcttccgccc 795
Db 1103 GAGGTGAAGGCTCGCTACAGCGCTACATCTCAAGCGGCGAGCTACGTTTCCGGCTC 1162
QY 796 agctgtactccctcggggcagtgggccatcggctatgtgagctcagatggcagcactcgt 855
Db 1163 TCTTCGACGCGATTGGGGCCATGGGCCATCGGTACGTAACTGCCGAGGAGAGATTCTG 1222
QY 856 cagaccatccctgcacaaacccctgtcccaggtgtcctgagggagacagaggagcgc 915
Db 1223 CAGACAATCCCTCAGACAAGTCCCTGTCGACAGGCGCTGCTCGATGCCCATCCGAGGGC 1282

```



```

VERSION      U26711.1  GI:862408
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 3241)
AUTHORS      Keane.M.M., Rivero-Lezcano.O.M., Mitchell.J.A., Robbins.K.C. and
             Lipkowitz.S.
TITLE        Cloning and characterization of cbl-b: a SH3 binding protein with
             homology to the c-cbl proto-oncogene
JOURNAL       Oncogene 10 (12), 2367-2377 (1995)
MEDLINE      95303504
AUTHORS      2 (bases 1 to 3241)
             Lipkowitz.S., Keane.M.M. and Mitchell.J.A.
TITLE        Direct Submission
JOURNAL       Submitted (10-MAR-1995) Stan Lipkowitz, Navy Medical Oncology
             Branch, National Cancer Institute, Bldg 8, Rm 5101, Bethesda Naval
             Hospital, Bethesda, MD 20889, USA
FEATURES     Location/Qualifiers
             1..3241
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /cell_line="breast cancer cell line 2R75-1"
             /map="3q"
             /chromosome="3"
             323..2755
             /note="binds SH3 proteins; similar to c-cbl proto-oncogene
             product, Swiss-Prot Accession Number P22681"
             /codon_start=1
             /product="cbl-b truncated form 1"
             /protein_id="AAB09292.1"
             /db_xref="GI:862409"
             /translation="MANSMGRNPGSGRGRILGIDAIQDVGPPKQAAARRR
             TVEATKMDKVRQKQKPLQKNSPPYILDLPITYQHLRLILSLSDNOKLAQLS
             ENYFIYIDSLMKSKRAIRLFKEGKERMVEQSDRRNLKLSIFSHMLAEIKAI
             FPNQFQDNFRITKADAAEFKRFKFDKTVPKVFRCLHEVHIFSSSLAMALKS
             TIDLNDYISVEFDITRLFPWGSILRNWFLAVHPGYMAFTYDEVKARLOKY
             STRGSVIFRLSTRLGOWALGYVTDGNILQIPENKPLFOALIDGSEGVLYPDG
             RSYNPDLTGCEPTPHDIKVTQELICEMGSTFQLCKICAEKXDKVILPFCGLHM
             CISLAWQSDGCGPCFCEIKGTEPIIVDFDPDESGRCCSIDFEGMPLDLD
             DDDRESLMNLANVKCTDRNSVTPSPGSLAQRKQPDPLQIPLSLPVP
             PRDLIQGVIRSPCGSGTSPKSCFWKQKPLPAPPPLRDPPLPPPPPPPIIP
             PNDLSHIIHVSVPSPDPMPLEAWCPDRVFTNQLVGLRGSPKPGTASSN
             VNGHRSVGLSDPVLMMRHRDLPLLEGAKVFSNHLGSEYDVPRLSPPPVTTLLP
             SIKCTGPLANSLEKTDPEDEDDDEYKIPSHPSINSOPSHCHNKPVPVSCDNGH
             CMLNGTHGPSEKSNIPDLSIYLKGVDFDSADPVLPLPAPPPTRDNPKHSSLNRT
             PSDIDLILPLLG"
             668..679
             /note="encodes nuclear localization signal"
             1439..1558
             /note="encodes ring finger"
             2751
             /note="alternative splice at nucleotide 2751 truncates
             protein before leucine zipper"
BASE COUNT   907 a 758 c 720 g 856 t
ORIGIN
Query Match      22.8% Score 323.6; DB 91; Length 3241;
Best Local Similarity 59.48; Pred No. 3.4e-48;
Matches 548; Conservative 0; Mismatches 374; Indels 0; Gaps 0;
Qy 351 caggcagcagctggcagctgcatcatcttcagccacatgcacgcagcagctgcacgc 410
Db 739 CAGAGCAAACTCTCAGCAAACTGTCCTATCTTCAGTCACATGCTGCAGAAATCAAGC 798
Qy 411 actctctccggggcagaagctacttgacacatgtacacgtaccacgaagcccccga 470
Db* 799 AATCTTTTCCCAATGGTCAATTCAGGAGATACTTTCGATACCAAAAGCAGATGTCG 858
Qy 471 cacctctcggagggaagcttcggagccgggtgtgtctgcctctgggctgagttgagtc 530

```

```

Db 859 TGAATTCGGAGAAAGTTTTTTTGGAGACAAAACATATCGTACCATGGAAGTATTGAGACA 918
Qy 531 cctcctggggcaccctgcaccctgtggaccaggtgtcacagccctggcctgtgcgcaccac 590
Db 919 GTCCCTTCATAGGTCCACACAGATTAGCTCTAGCTGGAGCAANTGGCTCTAAATCAAC 978
Qy 591 cattgacctcucctgcagcgggcaagtgtccatctcttcgagttgcagctgttcaccaggt 650
Db 979 AATTGATTAACTTGAATGATACATTTTCACTTTTGAATTTGATTTTACCAGGCT 1038
Qy 651 ctctcagccatggcaccactcctcaagactggcagctcctcggcagtcacacccacgg 710
Db 1039 GTTTCAGCCTTGGGGCTCTAATTTGGGGAATTTGGAATTTCTTAGCTGTGACACATCCAGG 1098
Qy 711 ctacatggcctctcctcaccctatgatgaggtccaaagaggtctgcaggcctgcaggagcaa 770
Db 1099 TTACATGGCATTTCTCACATATGATGAAGTTAAAGCAGACATACAGAAATATAGCACC 1158
Qy 771 gccagcagttacatctctccggccacagctgtactgcctggggcagtggtggccatcggt 830
Db 1159 ACCGGGAAGCTATATTTTCCGGTTAAGTTGCACTCGATTGGGACAGTGGGGCATTTGGGCTA 1218
Qy 831 tgtgagctcagatggcagcatcctcgcagaccatcctcgcacacacccctgtcccaggt 890
Db 1219 TGTGACTGGGGATGGGAATATCTTACAGACCATACTCATACAGCCCTTATTTCAAGC 1278
Qy 891 gctcctgggggacagagagcaggtcttacctctaccacagatggaagaccacacacc 950
Db 1279 CCTGATTGATGGCAGCAGGGAAGGATTTTATCTTTATCCTGATGGGAGAGTTATAATCC 1338
Qy 951 aaacctgactgagctcggcaggcagaacccacagcagcagcagcagcagcagcagcagc 1010
Db 1339 TGATTTAATGATGATATGTGAACCTACCTACCTCATACATATAAAGTTACACAGGAACA 1398
Qy 1011 gctgcagctctactggggcagcagcagcagcagcagcagcagcagcagcagcagcagc 1070
Db 1399 ATATGAATTAATTTGAAATGGGCTCCACTTTTCAGCTCTGTAAGATTTGTGCAGAGAA 1458
Qy 1071 caacaagatgtgaagattgagcagcagcagcagcagcagcagcagcagcagcagcagc 1130
Db 1459 TGACAAAGATGTCAAGATTGAGCCTTGTGGGCATTGTGACCTTTGCTTACGGC 1518
Qy 1131 ctggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1190
Db 1519 ATGGCAGGAGTCGGATGGTGCAGGGCTGCCCTTTCTGCTGTGTAATAAAGGAACTGA 1578
Qy 1191 ggcctgagctatctaccagttccacagcagcagcagcagcagcagcagcagcagcagcag 1250
Db 1579 GCCCAATATCGTGGACCCCTTTGATCCAAAGATGAAGGCTCCAGGCTTTCAGCATCAT 1638
Qy 1251 tgaccaggaagcagggaggttg 1272
Db 1639 TGACCCCTTGGCATGCCGATG 1660

```

```

RESULT      8
HSU26712    3354 bp mRNA PRI 30-SEP-1996
LOCUS       Human cbl-b truncated form 2 lacking leucine zipper mRNA, complete
DEFINITION  cds.
ACCESSION   U26712
VERSION     U26712.1 GI:862410
KEYWORDS    human.
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3354)
AUTHORS     Keane.M.M., Rivero-Lezcano.O.M., Mitchell.J.A., Robbins.K.C. and
             Lipkowitz.S.
TITLE       Cloning and characterization of cbl-b: a SH3 binding protein with
             homology to the c-cbl proto-oncogene
JOURNAL     Oncogene 10 (12), 2367-2377 (1995)

```


95303504
2 (bases 1 to 3354)
Lipkowitz, S., Keane, M.M. and Mitchell, J.A.
Direct Submission
Submitted (10-MAY-1995) Stan Lipkowitz, Navy Medical Oncology
Branch, National Cancer Institute, Bldg 8, Rm 5101, Bethesda Naval
Hospital, Bethesda, MD 20889, USA
Location/Qualifiers
1..3354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="breast cancer cell line ZR75-1"
/map="3q"
/chromosome="3"
323..2635
/note="similar to c-bcl proto-oncogene product, Swiss-Prot
Accession Number P22681"
/codon_start=1
/product="cbl-b truncated form 2"
/protein_id="AAB09293.1"
/db_xref="GI:862411"
/translation="VANSNGENPGGRGNPKGRILGILDALDVGPPKQAAADR
TVETWKLMDKVRVQCPKQLKNSPPYILDLDTQVHLRLILSKYDDNQKLAQLS
ENEYKFIYDLSLMKSKRAIRLFKGRKMEYEQSDRRNRTKUSLISHMLAEIKAI
FNGQFQDNFRITKADAAEFWRKFEGDKTIYPVQVFRQCLHEVHQISSLEAKLKS
TDLACNDYISVFEDIFTRLPQPGSILRNWFLAVHPGYMAFLTYDEVKARLQKY
STKPGSYIFRLSCTRLGWAIGYVTDGNILOTIPHNKPLFQALIDGREGFYLYPDG
RYNPLDGLCSPDPDHTKVTQEOYELVCEMGSTFOLCKICAENDKVKTEPCGHLM
CTSLTAQESDQGCPCRCIKETPIIIVDPDPDEGSRCCSIIIDFGMMLDL
DDDEESLMNRLANVRKCDTRQNSPVTSPGSSPLAARRPQDPDLPIRLHSLPPVP
PRDLTKQIVASPGSPGTPSPKSVVRKQDKLPAPPPPLRPDPDPDPDPDPDP
PNRLSRTHVHSVSPDPMPLEAWCPVRDVTGNLQVGRLLGEGSPKPGITASSN
VNGRHSRVGSDSLVLRKRRHDLPLEGAKVFNHGLSEEDYDVPRLSPPPVITLLP
STKCTGFLANSLETRDVEDDEYKIPSSHPVSLNSQPSHCHNVRPPVRSQDNH
CMLNTHGSPSEKSNIPDLISYLKGYRI"
668..679
/note="encodes nuclear localization signal"
1439..2023
/note="encodes ring finger"
2619
/note="alternative splice at nucleotide 2619 truncates
protein before leucine zipper"
BASE COUNT 943 a 773 c 739 g 899 t
ORIGIN

Query Match 22.8%; Score 323.6; DB 91; Length 3354;
Best Local Similarity 59.4%; Pred. No. 3.3e-48;
Matches 548; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

Qy 351 caggcgacagctggccaaagctatcttcagccacatgcacgcagagctgcacgc 410
Db 739 CAGACGAAATCTCACAAAACGTCCCTTATCTTCAGTCACATGCTGCAGAAATCAAGC 798
Qy 411 actcttcccgggcgaagctactgtgacacatgtaccagctcaccaagcccccgcaca 470
Db 799 AATCTTCCCAANTGGTCAATTCAGGGAGATAACTTTCGTATCAGAAAGCAGATGCTGC 858
Qy 471 cactcttgaggaggaagcttcgagcccggtgtgtgctgccccctggcctggagttgagtc 530
Db 859 TGAATCTTGAGAGAAAGTTTITGGAGACAAAACATACTGTAACCATGGAAAGATATTCAGACA 918
Qy 531 cctctctgggacagctgcaccccttggaacagagctgcacagccctggccttggcacccac 590
Db 919 GTGCGCTTCATAGTGCACAGATTAGCTCTAGCCCTGGAAGCAATGCTCTAAAATCAAC 978
Qy 591 cattgacctcacctgcagcgggacgctgtccatctctcagttcagcttcacccaggtct 650
Db 979 AATGATTTTACITGCAATATACATTTTCAGTTTITGATTTGATTTTATTTTACCAGGCT 1038
Qy 651 ctttcagccatggccacacatctctcaagaactcggcagctcctctggcagtcacacccagg 710
Db 1039 GTTTCAGCCTTGGGGCTCTATTTTTCGGGAATTTGGAATTTTCTTACTGTGACATCCAGG 1098

Qy 711 ctacatcgcttctcactacatgatgaggtcccaagagcgtctgcagcgcttcgagggacaa 770
Db 1099 TTACATGCAATTTCTCACATATGATGAAGTTAAAGCAGCAGTACAGAAATATACACCAH 1158
Qy 771 gccaggaggttacatctctccggccacagctgtactcgcctggggcagtgggccatcgcta 830
Db 1159 ACCCGAAGCTATATTTTCGGTTAAGTTGCACTCGATGGACAGTGGCCATTGGCTA 1218
Qy 831 tctgagctcagatggcagcattcctgcagaccatccctgcacacacacacacacccctgtccaggt 890
Db 1219 TGTGACTGGGATGGGATATATCTTACAGACCATACCTACATAACAGGCCCTTATTTCAGC 1278
Qy 891 gctcctggagggacagagggcgtctcactctaccctaccagatggaaagacccacacccc 950
Db 1279 CTTGATTGATGGCAGCAGGGAAGGATTTTATCTTTATCTTATCTTGATGGAGGATTATAATCC 1338
Qy 951 agacctgactgagctggccagggcaggaaccccccagcagcagcagcagcagcagcagcagc 1010
Db 1339 TGATTTAATCTGATTTGTGAATGGCTTCCACTTTTCAGCTCTGTAAGATTGTGCGAGAGAA 1398
Qy 1011 gctcagctctactggcggcagctgactccacatttgagctctgcaagatctgctgtgagag 1070
Db 1399 ATATGAATATTTGTGAATGGCTTCCACTTTTCAGCTCTGTAAGATTGTGCGAGAGAA 1458
Qy 1071 caacaagatgtgaagattgagccgctggggcagcactgctctgcagctgctgcctggctgc 1130
Db 1459 TGACAAAGATGTCAAGATTGAGCCTTGTGGCATTTGATGTGCACCTCTTGCCTTACGGC 1518
Qy 1131 ctggcagcactcggacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1190
Db 1519 ATGGCAGAGTCGGATGTCAGGGCTGCCCTTTCTGCTGTGTAATAAAGGAATGA 1578
Qy 1191 ggcctgagtgatctaccagttccacggtccagctactgctgagcagcagcagcagcagcag 1250
Db 1579 GCCCATATCTGGACCCCTTTGATCCAGAGATGAAGGCTCCAGGTTTTCAGCATCAT 1638
Qy 1251 tgaccaggaaggcagggaggtg 1272
Db 1639 TGACCCCTTGGCATGCCGATG 1660

RESULT 9
HSU26710
LOCUS HSU26710 3982 bp mRNA PRI 30-SEP-1996
DEFINITION Human cbl-b mRNA, complete cds.
ACCESSION U26710
VERSION U26710.1 GI:862406
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3982)
AUTHORS Keane, M.M., Rivero-Lezcano, O.M., Mitchell, J.A., Robbins, K.C. and Lipkowitz, S.
TITLE Cloning and characterization of cbl-b: a SH3 binding protein with homology to the c-bcl proto-oncogene
JOURNAL Oncogene 10: (12), 2367-2377 (1995)
MEDLINE 95303504
REFERENCE 2 (bases 1 to 3982)
AUTHORS Lipkowitz, S., Keane, M.M. and Mitchell, J.A.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-1995) Stan Lipkowitz, Navy Medical Oncology Branch, National Cancer Institute, Bldg 8, Rm 5101, Bethesda Naval Hospital, Bethesda, MD 20889, USA
FEATURES
source
1..3982
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="breast cancer cell line ZR75-1"
/map="3q"
/chromosome="3"

CDS 323 3271
 /note="interacts with SH3 proteins; similar to c-cbl
 proto-oncogene product, Swiss-Prot Accession Number
 P22681"
 /codon_start=1
 /product="cbl-b"
 /protein_id="AAB09291.1"
 /db_xref="GI:862407"
 /translation="MANSNGRNPGRGGRNPRKGRILGIDAIQAVGPPKQAAADR
 TVETKMLDKVVLQCNPKLQKNSPPKILDLPTDYQHLRLILSKYDDNOKLAOLS
 ENYFKIYIDSLMRKSKRAIRLEKEGERMEYEQSDRRNLKLSIFSHMLAEIKAI
 FPGQGDNDRIKADAAEFWRKFDKTIIVKVFROCLHEVHQIISLSLMAALKS
 TIDLTQNDYISVEFFIDFTLFPQWGLIRNNFLAVTHPGVMAFTYDEVKARLQKY
 STKPSVIFRLSCTRILGQWAGVTVGDGNILOTIPNKLFOALIDGSEGVLYPDG
 RSNPDLTGICETPHDHLKVTOEQYELICEMGSTFQICKICAENDKDKVIRPCEHLM
 CTSLTAWQSDGCGCFKCEIRTEPIIVDFPRDGRSCCSLIIDFGMPMLDLD
 DDDRESLMNRLANVKCTDRQNSPVTSPGSSPLAQRKFPDPLQIPHLSLPVP
 PRDLIQGIVRPGSPGSPKSPCMVKQDKPLPAPPPLRDPPLPPPPRPPPIIP
 PDLRLSRHIIHVESVSPRPPMPLEAWCPDRDVFNTQLYGCRLLGEGSPKPGITASSN
 VNGRHSRVDSPVLMKRRHRLPLEGAKVFNHLSGSEYDVVPRLSPPPVTVTLPL
 SIKCTGLANLSSEKTRDVEEDDEYKIPSSHPVSLNSQPSHCNKPVPVSCDNGH
 CMLNGTHSPSSSEKSNIPDLSIYLGKDVDSADVPPLPAPPRTDRNPKHSSSNRT
 PØDIDLIPPLGSDADALPSPPLPPPPFARSLIEHSPGSSSRPSGQDLFLPS
 DPVDLASQGVPLPPAPRUPGENVKNRTSQDYDLPSCSDGSPAPRPPKPRRTA
 PEIHRKPHGPFAALENDVDAKIAMGEGAFEEVKRALEIAQNNVVEARSILREFAF
 PPVPSPLNL"
 668 679
 misc_feature
 /note="encodes nuclear localization signal"
 1439 1558
 misc_feature
 /note="encodes ring finger"
 3116 3223
 misc_feature
 /note="encodes leucine zipper"
 BASE COUNT 1093 a 969 c 877 g 1043 t
 ORIGIN

Query Match 22.8%; Score 323.6; DB 91; Length 3982;
 Best Local Similarity 59.4%; Pred. No. 3.1e-48;
 Matches 548; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

QY 351 caggcgacgtggccagctggccatctcattcagccacatccagcagcagctgcacgc 410
 DB 739 CAGACGAAATCTCAAAATGTCCTTATCTTCAGTCACATCTGCGAGAAATCAAGC 798
 QY 411 actttcccccgggcaagtactgtggacacatgtaccagctcaccaggcccccacca 470
 DB 799 AATCTTTCCCAATGTCATTCAGGAGATAACTTTCGTATCATCAAAAGCAGATGCTG 858
 QY 471 cactctcagagaaagtgcgagcccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 530
 DB 859 TGAATTCGAGAAAGTGTGAGACAAACTATCGTACCATGGAAGATATTCAGACA 918
 QY 531 cctctggggcactgcaccctgtggaaccaggctgcagagccctgctgtgcaccac 590
 DB 919 GTGCTTCATGAGTCCACCATATAGCTAGCTGGAAGCAATGGCTCTAAATCAAC 978
 QY 591 cattgaacctcactgcaggggacgtgtcattctcaggttcagcgtcttcacagggt 650
 DB 979 AATTGATTTAACTGCAATGATTACATTTGAGTTTGTGAAATTTGATTTTACAGGCT 1038
 QY 651 ctctcagccatggccacactcctcagaactggcagctcctggcagtcacacccaccag 710
 DB 1039 GTTTCAGCTTGGGGTCTATTTCGGAATTCGAAATTTCTAGCTGTCACATCCAGG 1098
 QY 711 ctacatggccttctcactcatatgataggtccaaagcgtctgcagccctcagaggaca 770
 DB 1099 TTACATGGCAATTTCTACATATGATGAAGTTAAAGCACCACATACAGAAATATAGCACC 1158
 QY 771 gccaggcagttacatctctcggcccgagctgtactcctcggggcagtcggccatcggtca 830
 DB 1159 ACCGGAGAGCTATATTTCCGGTTAAGTTCACCTCCATTCGATGGGACAGTGGCCATGGCTA 1218
 QY 831- tgtgagtcagatgagcagcatcctgagaccatccctgcacacacacccctgtccaggt 890

Db 1219 TGTGACTGGGGAAGGAATACTTTACAGACCATACCTCATACAGCCCTTATTTCAGC 1278
 QY 891 gtccttgaggagacagaggagcgttctacctcaccagatggaagacccacaccc 950
 Db 1279 CCTGATTGATGACACAGGGAAGGATTTTATCTTTATCTTGTGGGAGAGTTATATCC 1338
 QY 951 agacctgactgagctcggccggggcagaccccccagcagcgcacacgtgtcagggagca 1010
 Db 1339 TGATTTAACTGGATTATGTGAACCTTACACCTCATGACCATATAAAGTTACACAGGAACA 1398
 QY 1011 gctgcagctctactggccatggactccacatttgagctctgcaagatctctgctgagag 1070
 Db 1399 ATATGAATATATTGTGAATGGGCTCCACTTTTCAGCTCTGTAAAGATTGTGCAGAGAA 1458
 QY 1071 caacaagatgtgaagattgagccgtgcgggcaacctgtcgcagctgctgcctggctgc 1130
 Db 1459 TGACAAAGATGTCAGATTTGAGCCTTGTGGCATTTGATGTGCACCTCTTGCCTTACGGC 1518
 QY 1131 cggcagcactgcagacacccagcactgccccttctgcgcgtgcagatcaaggctggga 1190
 Db 1519 ATGGCAGAGTCGGATGTCAGGCTGCCCTTCTGCGTTGTGAATAAAGGAAGTGA 1578
 QY 1191 ggcctgagatctaccagttccacggtcaggtcagctactgctgagactcagggaacacag 1250
 Db 1579 GCCCATATCTGGACCCCTTTGATCCAAAGATGAAGGCTCCAGGTGTTTCGCACATCAT 1638
 QY 1251 tgaccaggaagcaggaggttg 1272
 Db 1639 TGACCCCTTGGCATGCCGATG 1660

RESULT 10
 MMCCBL 2808 bp mRNA ROD 12-DEC-1992
 LOCUS Mouse mRNA for c-cbl proto-oncogene.
 DEFINITION X57111
 ACCESSION X57111.1 GI:50314
 VERSION cbl oncogene; nuclear protein; oncogene cellular.
 KEYWORDS cbl oncogene; nuclear protein; oncogene cellular.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 2808)
 AUTHORS Direct Submission
 TITLE Submitted (02-JAN-1991), W.Y. Langdon, INSTITUTE OF MEDICAL &
 JOURNAL VETERINARY SCIENCE, IMVS, DIVISION OF HUMAN IMMUNOLOGY, BOX 14
 RUNDEL MALL POST OFFICE, ADELAIDE SA 5000, AUSTRALIA
 REFERENCE 2 (bases 1 to 2808)
 AUTHORS Langdon, W.Y.
 TITLE Direct Submission
 JOURNAL Submitted (01-JAN-1991), Langdon W.Y., Institute of Medical and
 Veterinary Science, IMVS, Division of Human Immunology, Box 14
 Rundel Mall Post Office, Adelaide SA 5000, Australia
 REFERENCE 3 (bases 1 to 2808)
 AUTHORS Blake, R.J., Shapiro, M., Morse, H.C. III, and Langdon, W.Y.
 TITLE The sequences of the human and mouse c-cbl proto-oncogenes show
 v-cbl was generated by a large truncation encompassing a
 proline-rich domain and a leucine zipper-like motif
 Oncogene 6 (4), 653-657 (1991)
 JOURNAL 91232862
 MEDLINE
 REFERENCE 4 (bases 1 to 2808)
 AUTHORS Blake, R.J. and Langdon, W.Y.
 TITLE A rearrangement of the c-cbl proto-oncogene in HUT78 T-lymphoma
 cells results in a truncated protein
 Oncogene 7 (4), 757-762 (1992)
 JOURNAL 92228506
 MEDLINE
 FEATURES Location/Qualifiers
 source 1..2808
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 mRNA 1..2808

gene /gene="c-cbl"
evidence-experimental
1. 2808
/gene="c-cbl"
97.2787
CDS /gene="c-cbl"
/codon_start=1
/product="c-cbl protein"
/protein_id="CAA40394.1"
/db_xref="GI:50315"
/db_xref="SWISS-PROT:P22682"
/translation="MAGNYKSSGAGGSGGSGAGSLIGLMKDAFOPHHHHLSPH
PPCTVDRKMKVCKWLMKVVRLCONFLAKNPPYILDLPTDYLHRTVLSRYEG
KMETLGENEYFRVEMENLKTQIISLFEGKEMYEENSQPRNLTSLIFSHML
AELKGFPSGLFOGDTFRITKADAAEFWRKAFGEKTIYPWKSFRQALHEVHPISGLE
AMALKSIDLTCDNYISVFEDIFRLFPQWSLLRNWNSLAVTHPGYMAFLYDEVK
ARLOKFTKPGSIFRLSCTRGOWAIGYVADNLIQTIPHNKLPQALIDGFEQF
VLPDGRNDRJLGLCEPPODHIKVTOICAENDKVKIEPCGHLMTSCLASWQES
EQGQFCRCBIKETEIVDPDRSGSLLRQAGAFSPNDDDDRRADSLFM
MKELAGKVRSPSPFMAPOASLPVPPRLDLQAPVASTSVLGTSKASGSL
HKDPLIPPTLRPLPPPPREDREYSVGAETRPQRPPLCTPGDGRDLKPPVFSR
PGDLSWLSRTIPKVPATPNPGDWNGLNRHLSLPSLSQMEPRADYPLGTSFL
DTSMTNNSPVAGPESHPKIPSSANAIYSLAARPLMPKLPFGQSGSEEDTEYM
TPTSRPVQKPEKRLPLEAOTOSRACDQOIDSCTYEAMYTIQSOALSVAENSAG
EGNLATHTSGPSENEEDGYDPKPPVAVLARETLDISNASSFGWLSLDGDP
TNFEGSOVPERPKPFRINRSEKASSYQGGGATANPVATPASPQLSEIERLMS
QGYSDIQKALVLAHNMIEMAKNILLREFVSISSPAHVAT"
BASE COUNT 691 a 796 c 693 g 628 t
ORIGIN

Query Match 19.0%; Score 270.6; DB 12; Length 2808;
Best Local Similarity 57.8%; Pred. No. 5.6e-39;
Matches 535; Conservative 0; Mismatches 339; Indels 51; Gaps 1;

QY 291 ggtcctccccgggcccgaagagtgccaaacgacagctcttccgggggggctccagact 350
DB 471 GCAGACTATCAGCCTCTTCAAGGAGGAGAAAGAGGATGATGAGAGATTCACGCC 530
QY 351 caggcagagctggcgaagctggcgcacatcttcgacacacgacgacgagctcagc 410
DB 531 TAGGCGAAACCTGACCAATATTCCTGATCTTCAGCCACATGCTGGGAGCACTAAAAG 590
QY 411 actctccccgggcaaaagtactgtggacacatgtaccagctccacaaagcccccgc 470
DB 591 CATCTTCCGACGGACTCTTCAAGGAGACACTTTCGGATATTAAGCTGATGCTGC 650
QY 471 cactcttgagggaagtggcggagccgggtgtgtgtgcctctgggctgggttgatg 530
DB 651 CGAATTTTGAGAAAAGCTTTTGGTGAAAAGACGATAGTCCCGTGGAGAGCTTCGACA 710
QY 531 cctcctgggcactgcacacctgtggaacagagctgcacagccgtgcttcgcacac 590
DB 711 GGCCTTGCATGAGTAGTCATCCCATCAGTCTTGGCGTGGAGGCCATGCGCTGAATCCAC 770
QY 591 cattgacctcactgcagcgggcagctgtccatcttcgagttcgagcttccaccaggt 650
DB 771 TATTGATCTGACCTGCAATGATTATTTCTGCTTTGAATTGATATTTTACACGGGT 830
QY 651 ctttcagcatggcacaactcctcgaagactggcagctccttcggcagtcacaacccagg 710
DB 831 GTTTCAGCCTGCTCTCTTCTGCTAGAAATTGGAACACGCTTGTGTAACACCCCTGG 890
QY 711 ctacatggcctctcctcacctatagagtcacaagagctctgcagggcctgcaggacaa 770
DB 891 TTACATGGCTTCTGACATACGATGAGTGAAGCGCCCTGCGAGAGCTTATCCACAA 950
QY 771 gccaggcagttacattctccggcccagctgtactgcctcggggcagtgggccatggcta 830
DB 951 ACCTGGCAGTTACATTTTCGGCTGAGCTGATCTCGTTTGGGTGAGTGGGCTATTGGGTA 1010
QY 831 tttgagctcagatggcagacatcctgcagaccatcctgcacaaacccctctccaggt 890

Db 1011 TGTTACTGCGATGGGAACAATCTGCAGACAATCCACACAATAAACCCGCTCTTCCAAAGC 1070
QY 891 gctcctggaggacagaagcgtctctacotctcaccagatgggaagacccacacccc 950
Db 1071 ACTGATGATGGCTTCAAGGAAGGCTTATTGTTTCTGATGAGGAGAAATCAAAATCC 1130
QY 951 agacctgactgagctcggccagggcagaacccccagcagcgcacatccactgtctcagagagagca 1010
Db 1131 TGACCTGACAGGTTTATGTGAACCAACTCTCAAGATCATATCAAAAGTAAC----- 1181
QY 1011 gctgcagctctactggggccatgactccacatttgactctgcaagatctgtgctgagag 1070
Db 1182 -----CCAGATATGCTGAGAA 1199
QY 1071 caacaagagattgaagattgagccgtgcggggcaccctgctctgagctgtgctgctgctgc 1130
Db 1200 TGATAAGGATGTGAAGATTGAGCCCTGTGGACACCTCATGTGCACATCTCGCTCAGCTC 1259
QY 1131 ctggcagcactcggacagcagcagcaccctcctccctcctccgctgcgagatcaaggctgggaa 1190
Db 1260 GTGCAGGAATCAGAGGTCAGGGCTGTCTTTTCCGATGTAATCAAGGTAAGTACTGA 1319
QY 1191 ggcctgagtagtatccacagttccac 1215
Db 1320 GCCATCGTGTGATCCGTTTGAC 1344

RESULT 11
CASNS1
LOCUS
DEFINITION
Accession
Version
Keywords
Source
Organism
Reference
Authors
Title
Journal
Medline
Comment
Features
Source
CDS

CASNS1 2711 bp DNA VRL 02-AUG-1993
LOCUS Mouse Cas NS-1 retrovirus gag-onc fusion protein (v-cbl) gene, 3'
DEFINITION end.
Accession J04169
Version J04169.1, GI:323269
Keywords gag-onc fusion protein; viral oncogene.
Source CAS NS-1 retrovirus DNA, isolated from mouse (strain NFS/N)
previously infected with Cas-Br-M virus.
Organism Cas NS-1 murine leukemia virus
Viruses: Retrovirus; Retroviridae; Mammalian type C
retroviruses: 1-Mammalian type C virus group.
Reference 1 (bases 1 to 2711)
Authors Langdon,W.Y., Hartley,J.W., Klincken,S.P., Ruscetti,S.K. and
Morse,H.C.III.
Title v-cbl, an oncogene from a dual-recombinant murine retrovirus that
induces early B-lineage lymphomas
Journal Proc. Natl. Acad. Sci. U.S.A. 86, 1168-1172 (1989)
Medline 89145204
Comment Draft entry and computer-readable sequence for [1] kindly submitted
by W.Y.Langdon, 02-MAR-1989. Cas NS-1 is an acutely transforming
murine retrovirus that induces pre-B and pro-B cell lymphomas. It
was generated from the ecotropic Cas-Br-M virus by sequential
recombinations with endogenous retroviral sequences and a cellular
oncogene. The position of the oncogene junction was determined to
be in the gag gene for p10 at a position that removed the terminal
24 gag-encoded amino acids. The gag-onc fusion protein appears to
be responsible for fibroblast and pre-B cell transformation. The
authors call this oncogene v-cbl for Casitas B-lineage lymphoma.
Features
1. .2711
Location/Qualifiers
/organism="Cas NS-1 murine leukemia virus"
/db_xref="taxon:11793"
/c1..1509
/notes="gag-onc fusion protein"
/codon_start=2
/protein_id="AAA42885.1"
/db_xref="GI:323270"
/translation="LELKLDSOVHVDPEDPGQETNVMSTFWSAPDPIGRKLERLE
DLKSLGLDYRAEAEIFNKRREPEREIKETKEKERRAEDEQREKDRRRH
REMSKLATVIGRODRGRRRRFOLDKQCYKQETGHWAAAGGAGGRRGSP
GSIPSLAAERAPDPLAMAGNVKXSGAGGGGGGAGGLGMDADQPHHHHEL
SPHPPTDKKVKWKVLCQNPVALKNSPPYILDLPTDYLHRTVLSR
YEGKMETLGENEYFRVEMENLKTQIISLFEGKEMYEENSQPRNLTSLIF
HMLAELKGIFFPSGLFOGDTFRITKADAAEFWRKAFGEKTIYPWKSFRQALHEVHPIS

ORIGIN

```

/db_xref="GI:895767"
/db_xref="SPIRENBL.Q19019"
/translation="MGSINIFHRIIFRVNGTGNNAARFVSTNNSTEAATLSRAPVES
TVSLFPGASEMPFGCEDRFLFKACKFMQVVKSCSPRLNKNSPPFIDITIP
DAYTHMLFTQNDLQNDYLIKLEISMINKCKIILKLFKSAIYNDQSEERKLT
KWSLTSKMLFEIKALFPESGIVTEDFRRTKKEASEWSHFTKNIVPWSFTFALE
KHGSLTIGKMAEALKATLIDSCDDISINFEFVDFRFLFPFKTLKNWOTLTAPG
YCAFLYDIEGAELEKLTIDGSGTIFRLSTREGQAKRGIVADGKLIYIPQNKSLI
QALHEGKSGFYIYPNGROQDNLKMLDPVADRVQVTEQVELCEMGTTELCKLI
CDNENKIKIEPGCHLACKLANWQDSGGTCTPCFVIEIKGNRVIIDRKPTKLI
ETIEKANVAABEKLLSLVPDVPRTVVSOCOSLLHDASNISPSVDELPLVPPPLP
KALGTDLTLNSSOTSSSYNIKELENVETSGEALQVNVNQRAPSTQAPLPRLSAS
EHOFPHPHTNNSERE"
701.a 487 c 429 g 613 t
BASE COUNT
ORIGIN

```

Query Match	14.0%	Score 199.4	DB 33	Length 2230
Best Local Similarity	53.5%	Pred. No. 3.8e-28		
Matches 466	Conservative	0	Mismatches 396	Indels 9
Gaps				
Qy	353	ggcgacagctggccaagctggccatctcttcagccacatgcagcgagagctgcacgac	412	
Db	531	GACGGAACTTACGAAATGTCATACATTTTCACATATGCTTTTCGAGATTAAAGCAT	590	
Qy	413	tcttccocggggcaaaagtactctgtggacacatgtaccagctccacaagcccccocccaca	472	
Db	591	TATTTCCGGAGGTTATCTATATTGAAGACCGGTTTCGGATGACAAGAAGAAAGCCGAAA	650	
Qy	473	ctctctggagggaaagtgtgcggagcccggtgtgtgtcgtccctggcgtagtttgagtcgc	532	
Db	651	GCTTTTGGAGTCATCATTTTCAAAAAAACAATGTACCTGTGTCAACATTTTCTACTG	710	
Qy	533	tcctggggcacctggca---ccctgtggaaaccagctgcacagccctggccttgcgcaca	589	
Db	711	CATTAGAAAACCCATCGGATCAAGTAGAGAAAATCGAAGCAGCCGAAATTAAGTGTA	770	
Qy	590	ccattgacctcaactgcagcgggcacgtgtccatctcttcagtttcagcttcctaccaggc	649	
Db	771	CGATAGACTTCAGCGGAGATGATTTTATTTCGAATTTTGAGTTTGATGTGTTTACAAGGT	830	
Qy	650	tctttcgcgaatggccaacacactctccaagaactggcagctcctggcagtoaacacccag	709	
Db	831	TATTCTACCCCTTCAAAACACTGTATCAAAAATTGGCAAACTACCAACCGCCATCCG	890	
Qy	710	gtcatatggccttctcactatgatgaggttccaagagcgtctgcagcctgcagggaca	769	
Db	891	GATACTGTGCATTTCTCATATCGATGAGGTCAAAAACGGTTAGAAAAATTAACGAAA	950	
Qy	770	agccaggcagttacatcttccggccacgtctactctgcctgcctggggcagtgggccatcgct	829	
Db	951	AACCTGGAAGCTATACATCTTCGGGTTTATCATGCACACGTCCTGGACAATGGCAATAGAT	1010	
Qy	830	atrgtgcagctcagatggcagcatcctgcagacatccctgccecaaaacccctgtcccagg	889	
Db	1011	ACGTAGCTCCGGATGGAAGAATTTATCAGACAATACCACAGATAAAGTTTGATTCAG	1070	
Qy	890	tgcctctggagggacagaagcagcgtctctacctctaccagatggaaagaccacaacc	949	
Db	1071	CACATACATGAAGCCATAAAGAAGGATTTTATTTTACCCGAACGTTAGAGATCAAGATA	1130	
Qy	950	cagacctgactgcagctcggccagcaggcagaaccccagcagcgcatcacgtgtcagaggac	1009	
Db	1131	TTAACTTTATCCAAATTTGATGATGTGCCAACGCGGACAGATGGCAATGACCTGAAC	1190	
Qy	1010	agctgcagctctactggggccatggactccacatttgactctctgcagatcctgtgtgaga	1069	
Db	1191	AATACGAGTTTGTATGTGAGATGGCCACAACATTCGAGTTGTGCAAAATTTGTGACGATA	1250	
Qy	1070	gcaacaagga-tgtgaagattgagccggtgcgggacactctctgcagctgcctgcctggctg	1129	
Db	1251	ACGAGAACAATCAAAATTCGAGCATGTGACATTTGCTCTGCCAAAATTTTGGCTTA	1310	

RESULT 13

CEDNASLI1 2230 bp DNA INV 15-NOV-1995
 LOCUS
 DEFINITION C.elegans DNA for sli-1 gene.
 ACCESSION X89223
 VERSION X89223.1 GI:895766
 KEYWORDS negative regulator; sli-1 gene.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 2230)
 AUTHORS Yoon,C.H., Lee,J., Jongeward,G.D. and Sternberg,P.W.
 TITLE Similarity of sli-1, a regulator of vulval development in C.
 elegans, to the mammalian proto-oncogene c-bi
 JOURNAL Science 269 (5227), 1102-1105 (1995)
 MEDLINE 95381043
 REFERENCE 2 (bases 1 to 2230)
 AUTHORS Yoon,C.H.
 TITLE Direct Submission
 JOURNAL Submitted (28-JUN-1995) C.H. Yoon, CALIFORNIA INSTITUTE OF
 TECHNOLOGY, Div. Of Biology, Pasadena, CA91125, USA
 FEATURES
 source
 1..2230
 location/Qualifiers
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:5239"
 /chromosome="X"
 /clone_lib="Barstead"
 /map="-20 to -24.50 map units"
 65..1813
 /gene="sli-1"
 65..1813
 /gene="sli-1"
 /function="negative regulator of LET-23-mediated signaling"
 /codon_start=1
 /protein_id="CAA61507.1"
 gene
 CDS

90550	91801: contig of 1252 bp in length
91802	91901: gap of unknown length
91902	93127: contig of 1226 bp in length
93128	93227: gap of unknown length
93228	94440: contig of 1213 bp in length
94441	94540: gap of unknown length
94541	95730: contig of 1190 bp in length
95731	95830: gap of unknown length
95831	97015: contig of 1185 bp in length
97016	97115: gap of unknown length
97116	98210: contig of 1095 bp in length
98211	98310: gap of unknown length
98311	99370: contig of 1060 bp in length
99371	99470: gap of unknown length
99471	100529: contig of 1059 bp in length
100530	100629: gap of unknown length
100630	101680: contig of 1051 bp in length
101681	101780: gap of unknown length
101781	102825: contig of 1045 bp in length
102826	102925: gap of unknown length
102926	103948: contig of 1023 bp in length
103949	104048: gap of unknown length
104049	105054: contig of 1006 bp in length
105055	105154: gap of unknown length
105155	106153: contig of 999 bp in length
106154	106253: gap of unknown length
106254	107244: contig of 991 bp in length
107245	107344: gap of unknown length
107345	108328: contig of 984 bp in length
108329	108428: gap of unknown length
108429	109408: contig of 980 bp in length
109409	109508: gap of unknown length
109509	110479: contig of 971 bp in length
110480	110579: gap of unknown length
110580	111536: contig of 957 bp in length
111537	111636: gap of unknown length
111638	112557: contig of 921 bp in length
112558	112657: gap of unknown length
112658	113571: contig of 914 bp in length
113571	113671: gap of unknown length
113672	114559: contig of 888 bp in length
114560	114659: gap of unknown length
114660	115542: contig of 883 bp in length
115543	115642: gap of unknown length
115643	116516: contig of 874 bp in length
116517	116616: gap of unknown length
116617	117461: contig of 845 bp in length
117462	117561: gap of unknown length
117562	118402: contig of 841 bp in length
118403	118503: gap of unknown length
118503	119341: contig of 839 bp in length
119342	119441: gap of unknown length
119442	120275: contig of 838 bp in length
120280	120379: gap of unknown length
120380	121210: contig of 831 bp in length
121211	121310: gap of unknown length
121311	122141: contig of 831 bp in length

```
Query Match          12.6%; Score 179.2; DB 39; Length 129226;
Best Local Similarity 65.5%; Pred. No. 3.4e-23;
Matches 262; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
```

505 tgcctgcctcgtgctgagtttgagtcctcctcctcctggtggcacccctgtggaccaggct 565

Dbp 122325 TGGTTCCTCGGCGAGGCTTAACAAAGTACATCCATAATCTCCGGCC 122384

566 gcacagccctggccttgcgaccaccattgacctcacctgcagggggcacatgtccatct 625

Db 122385 TGGAGGCCATGGCCCTAAGGCCACTATCGATCTTACCTGCACGACTTCATTTCCAACT 122444

626 t cga gtt cga cgt ctt c a c c a g g t c t t t c a g c c a t c g c c a a c a c t c c t c a g a a c t q q c 685

bp 12245 TCGAGTTCGACGCTTTCACAGCCTCTTCCAGCCTTGGGTGACCTCGTACGCAACTGGC 122504

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 06:02:31 ; Search time 79.91 seconds
(without alignments)
2977.917 Million cell updates/sec

Title: US-09-434-708-1
Perfect score: 1547
Sequence: 1 cggaggtccatggtctt.....caaaaaaaaaaaaaaaaa 1547

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 73620496 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/5C.COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/5D.COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/6.COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
- 7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	340	22.0	3090	6	PCT-US93-06251-7
2	46.4	3.0	9960	5	US-08-822-586-46
3	42.2	2.7	1593	5	US-09-320-878-23
4	41.6	2.7	3292	5	US-09-320-878-22
5	40	2.6	1701	5	US-09-357-072-1
6	40	2.6	1701	6	PCT-US95-16342-1
7	40	2.6	1701	6	PCT-US96-10521-1
8	38.4	2.5	4692	3	US-08-916-917-1
9	38.4	2.5	4692	3	US-08-972-631-1
10	38.4	2.5	4592	3	US-08-972-639-1
11	38.4	2.5	4592	3	US-08-972-630-1
12	38.4	2.5	4692	3	US-08-672-211-1
13	38.4	2.5	4692	3	US-09-225-170-1
14	38.2	2.5	954	5	US-08-680-506-5
15	38.2	2.5	1347	5	US-08-680-506-8
16	38.2	2.5	1587	5	US-08-680-506-6
17	38.2	2.5	2233	5	US-08-680-506-4
18	38.2	2.5	5027	5	US-08-680-506-2
19	38	2.5	2150	5	US-09-263-083-1
20	38	2.5	5392	5	US-08-403-852D-1
21	38	2.5	5392	5	US-08-510-645B-1
22	38	2.5	15378	5	US-08-785-420-1
23	38	2.5	44377	3	US-08-804-227C-7
24	38	2.5	44377	4	US-08-804-198-1
25	37.4	2.4	53526	5	US-08-658-136-2
26	37.4	2.4	53577	5	US-08-658-136-1

27 37.2 2.4 3596 3 US-08-779-801-5 Sequence 5, Appli
28 37.2 2.4 3632 3 US-08-779-801-3 Sequence 3, Appli
29 37.2 2.4 3632 3 US-08-779-801-4 Sequence 4, Appli
30 36.8 2.4 1174 5 US-09-034-985-1 Sequence 1, Appli
31 36.8 2.4 3293 4 US-08-442-809A-75 Sequence 75, Appli
32 36.6 2.4 4451 5 US-08-717-294-42 Sequence 42, Appli
33 36.4 2.4 9723 5 US-08-083-590A-21 Sequence 21, Appli
34 36.4 2.4 9723 5 US-08-532-384-21 Sequence 21, Appli
35 36 2.3 1312 1 US-08-305-506A-1 Sequence 1, Appli
36 36 2.3 1312 6 PCT-US94-02389-1 Sequence 1, Appli
37 36 2.3 2558 2 US-08-477-476-1 Sequence 1, Appli
38 36 2.3 2558 4 US-08-893-210-1 Sequence 1, Appli
39 36 2.3 2558 6 PCT-US96-07354-1 Sequence 1, Appli
40 36 2.3 152331 5 US-09-128-155-16 Sequence 16, Appli
41 35.8 2.3 2286 5 US-09-176-657-4 Sequence 4, Appli
42 35.8 2.3 5996 5 US-09-528-934-33 Sequence 33, Appli
43 35.6 2.3 1560 4 US-08-709-959A-2 Sequence 2, Appli
44 35.6 2.3 2030 4 US-08-530-950-1 Sequence 1, Appli
45 35.6 2.3 2277 2 US-08-676-967-5 Sequence 5, Appli

ALIGNMENTS

RESULT 1
PCT-US93-06251-7
; Sequence 7, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Wickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06251
; FILING DATE: 19930630
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3090 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US93-06251-7

Query Match 22.0%; Score 340; DB 6; Length 3090;
Best Local Similarity 62.2%; Pred. No. 1.8e-69;
Matches 535; Conservative 0; Mismatches 325; Indels 0; Gaps 0;

QY 364 agggagcagctggccagctggccatcatttcagccacatgcacgagctgcacgca 423

DB 590 AGGCGAACCTAACCAAACTGTCCTCATCTTCAGCCACATGCTGGCAGAACTAAAGGA 649

QY 424 ctcttccccggggaagtagtctgtgacacatgtacacagctaccagagcccccccccac 483
|||||
Db 650 ATCTTCCAAAGTGGAGCTTTTCAGGAGACACATTTTCGGATTACTAAAGCAGATGTGG 709
QY 484 accttctggagggaagattcggagagcccggtgtctgccccctggcctgagtttgatcc 543
|||||
Db 710 GAATTTTGGAGAAAGCTTTTGGGAAAAGACAAATAGTCCCTTGGAGAGCTTTCGACAG 769
QY 544 ctctctgggcccctgcacccctctggaaaccaggtgtcacagccctggccttgccacacc 603
|||||
Db 770 GCTCTACATGAATGTCATCCATCAGTCTTCTGGCTGGAGGCCATGCTCTGAATCCACT 829
QY 604 attgacctaccctgcagggcggtgtccatctcttctgagtttgcacattcttccacaggtcc 663
|||||
Db 830 ATTGATCTGACCTGCAATGATATATTTTCGGTTTTCGATTGACATCTTTACCCGACTC 889
QY 664 ttccagccatgggcaacacccctcaagaactgcagctctctgcactcaaccacccagcc 723
|||||
Db 890 TTTGCGCCCTGGTCTTTTGTCTCAGGAATGGAACAGCTTCTGTACTCATCTCTGCG 949
QY 724 tacatggcctctctccactatgatgaggtccaaagagctgtcagagccctgcaggggacaag 783
|||||
Db 950 TACATGCTTTTGTGAGTATGACGAAGTGAAGCTCGGCTCCAGAAATTCATTCACAAA 1009
QY 784 ccagggaattacattctccggccagctgtactgcgctggggcagtgggccatcgctat 843
|||||
Db 1010 CCTGGCAGTTATATCTTCCGGCTGAGCTGTACTCGTCTGGGTGAGTGGGTATTTGGGTAT 1069
QY 844 gtgagctcagatggcagctctctgcagaccatccctgcacaaacccctgtcccagggtg 903
Db 1070 GTTACTGCTGATGGARACATCTCCAGACATCCCTCACAATAAACCCTCTTCCAGCA 1129
QY 904 ctcttgaggagacagaagggcgtcttacccttaaccatggaagagccacacccca 963
|||||
Db 1130 CTGATTGATGGCTTCAGGAAGGCTTCTATTGTTTCTGATGAGCAATACAGAAATCCT 1189
QY 964 gacctgtagctcggccaggcagaaacccagcagcagctcaccctgtcagagggagcag 1023
|||||
Db 1190 GATCTGACTGGCTTATGTGAACCACTCCCAAGACCATATCAAGTACCAGGAGCAAA 1249
QY 1024 ctgcagctctactggggccatggactccacatttgagctctgcaagatctgtgtgagagc 1083
Db 1250 TATCAATATATCTGAGATGGGCTCCACATTCCAATATGTAATAATGTGTGAAAT 1309
QY 1084 acaagatgtgaagattgagcctgcggcagcctctctgcagctgtcagctgctgctgctgc 1143
Db 1310 GATAGGATGTAAAGATTGAGCCCTGTGACACCTCATGTGACATCCCTGTCTTACATCC 1369
QY 1144 tggcagcactcggacagccagcctgcccccttctgcgctgcagatcaaggctgggag 1203
Db 1370 TGGCAGGAATCAGAAAGTCCAGGCTGTCCTTCTGCGGATGTAATTAAGGTACTGAA 1429
QY 1204 gcccgtgastatctaccagtt 1223
Db 1430 CCATCGTGTGATGCGTT 1449

RESULT 2

US-08-822-586-46
; Sequence 46, Application US/08822586
; Patent No. 6015890
; GENERAL INFORMATION:
; APPLICANT: WILLIAM R. JACOBS, JR., JAMES M. MUSSEY AND
; APPLICANT: AMALIO TELENIT
; TITLE OF INVENTION: AN EMBAC OPERON OF MYCOBACTERIA AND
; TITLE OF INVENTION: MUTANTS THEREOF
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK

RESULT 3

US-09-320-878-23/c
; Sequence 23, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BEILACH, Melanie C.
; APPLICANT: BEILACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li

COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
MEDIUM TYPE: DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,586
FILING DATE: MARCH 20, 1997
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/437
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 9960
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
US-08-822-586-46

Query Match 3.0%; Score 46.4; DB 5; Length 9960;

Best Local Similarity 45.2%; Pred. No. 0.066;
Matches 170; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

QY 831 ggcacatggctatgtgagctcagatggcagcagctcctcagaccatccctgcccacaaacc 890
Db 2385 GCGGTGCGGCGCGCTGACCTCGCTGGGCTTTCGGGCGGACGCGATCCCTCGGATGTCTC 2444
QY 891 cctgtccagctgtctctggaggacagagagcgtcttaccctaccacagatggaaa 950
Db 2445 CCGGACCCCTCATGTGAGCAGCCCGGCGACGACAACTTCCCGACAGCAGCTCCGCGT 2504
QY 951 gaccacacacccagacctgactgagctcggcagggcagaaacccagcagcgatccacgt 1010
Db 2505 GGTACCGGACCGAGGTGCGGAGGAGGCGGCGACAGCTGCCGCGGCGATCAACGG 2564
QY 1011 gtcagagagagctgagctcactggtggtggtggtggtggtggtggtggtggtggtg 1070
Db 2565 ATCCGCGCGCGCTGCGCTGCGGTACGCGCTGAACCGCGCGACACAGCGCGGTGCTCGT 2624
QY 1071 ctgtgtgagtgcaacagaggtgtgaagattgagcgtggtgggacacgtgtctgtcagctg 1130
Db 2625 CGGTGCGGCGACACAGAGCGCGCGGTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2684
QY 1131 ctgctgtgctcctggcagcactgggacagccagacacacacacacacacacacacacac 1190
Db 2685 CGACGAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2744
QY 1191 caagggtgtgagggcc 1206
Db 2745 CGAGGTGCGAGTGGGCC 2760

Qy 224 ggggtcccggggctctctggggacttttctactcatct 259
||| | | | | ||| |
Db 422 qgaagaaagacatgtgtgcagcatttaacgtcatat 457

```

RESULT      6
PCP-US95-16542-1
; Sequence 1, Application PC/TUS9516542
; GENERAL INFORMATION:
; APPLICANT: YEDA RESEARCH AND DEVELOPMENT CO. LTD.
; APPLICANT: WEINWURZEL, Henry
; APPLICANT: WALLACH, David
; APPLICANT: BOLDIN, Mark
; APPLICANT: VARFOLOMEYEV, Eugene
; APPLICANT: METT, Igor
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS/AP01
; NUMBER OF INVENTION: RECEPTORS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004

```

```

Query Match      2.8%; Score 40; DB 6; Length 1701;
Best Local Similarity 49.1%; Pred. No. 1.2;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

44 agtgggaagagcccgccgacctggccggcgatctggagcgccctagaagac 103
242 AGCGCAAGCTGGACGGCGGTACAGAGGGCTAGACTCTTCCTGATGCTGGAGCAGA 301

104 aatcgctgcagcccccggtctccgtgagtcgcccttccgtggggaacctgctgccccaca 163

```

Db	302	ACGACCTGGAGCCCGGCACACGAGCTCTCTGGGAGCTGCTGCTTCCTCGGGGCC	361
QY	164	csqcgacgtctctcgagagtggtgccattctcggcgggcgscggcgagagcgsgcccg	223
Db	362	ACGACCTGCTGCGGGCGTCTGACGACTTCAGGCGGGGCGCGGGCGCGCGCTG	421
QY	224	gsggtcccgggggtctgtgggacttttctaactatct	259
Db	422	GGGAAGAAGACTGTGTCCAGCATTTTAAAGTCATAT	457

```

RESULT 7
PCT-US96-10521-1
; Sequence 1, Application PC/TUS9610521
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
; NUMBER OF SEQUENCES: 34
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10521
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: IL 114,615
; FILING DATE: 16-JUL-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: IL 114,986
; FILING DATE: 17-AUG-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: IL 115,319
; FILING DATE: 14-SEP-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: IL 116,588
; FILING DATE: 27-DEC-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: IL 117,932
; FILING DATE: 16-APR-1996
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..768
PCT-US96-10521-1

```

	Query Match	2.6%	Score 40;	DB 6;	Length 1701;
	Best Local Similarity	49.1%	Pred. No. 1.2;		
	Matches 106;	Conservative	0;	Mismatches 110;	Indels 0;
	Gaps				
QY	44	agtgggaagagcgcccgccctggcgccggcgagtcagagatctcagcgctagaagagc	103		
Db	242	AGGCCAAGCTGGAGCGCGGTGCAGAGGGGCTAGACCTCTTCGATGCTGTGGACAGA	301		
QY	104	aatcgctcgacccccggcgctcgtagtcccccttcgctcgggacctcgtcccccga	163		
Db	302	AGGACTGGAGCCCGGGCACACCGAGACTCCTCGGGAGACTGCTCGCTCCTCGGGGGCC	361		
QY	164	cagcgcgagctcttcagagagtgggccattctcgggcgccggcgagggcgcccg	223		
Db	362	ACGACTCTGTCGCGCGTTCAGACTTCGAGGCGGGGGCGGGCGGGCGCGCCCTG	421		

QY 224 ggggtccggcggtctgggaactttctactctct 259
|||||
Db 422 GGAAGAAGACCTGTGTGACGCAATTACGTCATAT 457
|||||
RESULT 8
US-08-916-917-1/c
; Sequence 1, Application US/08916917
; Patent No. 5856132
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len
; APPLICANT: Hawkins, Phillip Thomas
; APPLICANT: Braselmann, Sylvia
; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,917
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/672,211
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8549-0006-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4692 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-916-917-1

Query Match 2.5%; Score 38.4; DB 3; Length 4692;
Best Local Similarity 50.9%; Pred. No. 3.6;
Matches 116; Conservative 0; Mismatches 111; Indels 1; Gaps 1;
QY 16 gctctggcggtggcccgctgggggacagtggaagagcccgccctggcgcgga 75
Db 1833 GCTTGGGCTGGGACGACGAGCGGGAG-CCTTGGGCCCGGACGGGGCAGGGTCGGGCTG 1775
QY 76 gtacgatgtctgcagcgccctagagagcaatgcgtcagcccccggtgtccgtgagtc 135
Db 1774 TCCAGGGGGCTGCAGAGGCTGCCGGCCCGACGAGGGCGGGCGCTGTCGGGGCTGCC 1715
QY 136 cctctgctgggacgtgtgccccacacagcgacgtcttctgagagtggtgcccattct 195
Db 1714 TCCAGGCTGCGCGAGTCCCTCCGACACCATCTGTGGTGTCTTTGAAGAGTTTATAG 1655
QY 196 cggcgggcgggcgggcgagggcgcccgggggtcccgcggtctctggg 243
Db 1654 ATCTGTGTAATCTCTCCCGGGCGGGGTGGCCCGGGCGCTCGTG 1607

RESULT 9
US-08-972-631-1/c
; Sequence 1, Application US/08972631
; Patent No. 5856133
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len
; APPLICANT: Hawkins, Phillip T.
; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 2730 Sand Hill Road
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,631
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/672,211
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,277
; REFERENCE/DOCKET NUMBER: 8549-0005-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)854-3660
; TELEFAX: (415)854-3694
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4692 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-972-631-1

Query Match 2.5%; Score 38.4; DB 3; Length 4692;
Best Local Similarity 50.9%; Pred. No. 3.6;
Matches 116; Conservative 0; Mismatches 111; Indels 1; Gaps 1;

QY 16 gctctggcggtggcccgctgggggacagtggaagagcccgccctggcgcgga 75
Db 1833 GCTTGGGCTGGGACGACGAGCGGGAG-CCTTGGGCCCGGACGGGGCAGGGTCGGGCTG 1775
QY 76 gtacgatgtctgcagcgccctagagagcaatgcgtcagcccccggtgtccgtgagtc 135
Db 1774 TCCAGGGGGCTGCAGAGGCTGCCGGCCCGACGAGGGCGGGCGCTGTCGGGGCTGCC 1715
QY 136 cctctgctgggacgtgtgccccacacagcgacgtcttctgagagtggtgcccattct 195
Db 1714 TCCAGGCTGCGCGAGTCCCTCCGACACCATCTGTGGTGTCTTTGAAGAGTTTATAG 1655
QY 196 cggcgggcgggcgggcgagggcgcccgggggtcccgcggtctctggg 243
Db 1654 ATCTGTGTAATCTCTCCCGGGCGGGGTGGCCCGGGCGCTCGTG 1607

RESULT 10
US-08-972-629-1/c
; Sequence 1, Application US/08972629
; Patent No. 5859201
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len

APPLICANT: Hawkins, Phillip T.
 TITLE OF INVENTION: G-BETA-GAMMA REGULATED
 PHOSPHATIDYLINOSITOL-3' KINASE
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 2730 Sand Hill Road
 CITY: Menlo Park
 STATE: California
 COUNTRY: USA
 ZIP: 94025
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/972,629
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/672,211
 FILING DATE: 27-JUN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Halluin, Albert P.
 REGISTRATION NUMBER: 25,277
 REFERENCE/DOCKET NUMBER: 8549-0005-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415)854-3660
 TELEFAX: (415)854-3694
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4692 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: cDNA
 US-08-972-629-1

Query Match 2.5%; Score 38.4; DB 3; Length 4692;
 Best Local Similarity 50.9%; Pred. No. 3.6;
 Matches 116; Conservative 0; Mismatches 111; Indels 1; Gaps 1;
 QY 16 gctctgctggtgccccgtggggcagacagtggaagagcccgccctggccgggca 75
 Db 1833 GCTTGGGCTGGGCGAGCGAGCGGAG-CCCTGGGCCCGGACGGGGCGAGTCCGGGCTG 1775
 QY 76 gtccagatgctgcagcgccctagaagacaaatgcgtgcagcccccggtgtccgtgagtc 135
 Db 1774 TCCAGGGGCTGCAGAGGCTGCCGCCCGACGAGGGGGCGGCTGTCCGGGCTGCC 1715
 QY 136 ccttcctgctggagacctgtgccccacacagcagctccttcagagagtgccccattct 195
 Db 1714 TCCAGGCTGCGGAGTCCCTCCGACACACCATCTGGCTGTGAAGAGTTTATAG 1655
 QY 196 cggcgggcgccggcggagcgcccgccgggggtcccgcggtctagg 243
 Db 1654 ATCCTGTGAACCTCTGCCGCGCGCGGTGGCCCGCGCTCGTG 1607

RESULT 11
 US-08-972-630-1/c
 Sequence 1, Application US/08972630
 Patent No. 5869271
 GENERAL INFORMATION:
 APPLICANT: Stephens, Len
 APPLICANT: Hawkins, Phillip T.
 TITLE OF INVENTION: G-BETA-GAMMA REGULATED
 PHOSPHATIDYLINOSITOL-3' KINASE
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds

STREET: 2730 Sand Hill Road
 CITY: Menlo Park
 STATE: California
 COUNTRY: USA
 ZIP: 94025
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/972,630
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/672,211
 FILING DATE: 27-JUN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Halluin, Albert P.
 REGISTRATION NUMBER: 25,277
 REFERENCE/DOCKET NUMBER: 8549-0005-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415)854-3660
 TELEFAX: (415)854-3694
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4692 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: cDNA
 US-08-972-630-1

Query Match 2.5%; Score 38.4; DB 3; Length 4692;
 Best Local Similarity 50.9%; Pred. No. 3.6;
 Matches 116; Conservative 0; Mismatches 111; Indels 1; Gaps 1;
 QY 16 gctctgctggtgccccgtggggcagacagtggaagagcccgccctggccgggca 75
 Db 1833 GCTTGGGCTGGGCGAGCGAGCGGAG-CCCTGGGCCCGGACGGGGCGAGTCCGGGCTG 1775
 QY 76 gtccagatgctgcagcgccctagaagacaaatgcgtgcagcccccggtgtccgtgagtc 135
 Db 1774 TCCAGGGGCTGCAGAGGCTGCCGCCCGACGAGGGGGCGGCTGTCCGGGCTGCC 1715
 QY 136 ccttcctgctggagacctgtgccccacacagcagctccttcagagagtgccccattct 195
 Db 1714 TCCAGGCTGCGGAGTCCCTCCGACACACCATCTGGCTGTGAAGAGTTTATAG 1655
 QY 196 cggcgggcgccggcggagcgcccgccgggggtcccgcggtctagg 243
 Db 1654 ATCCTGTGAACCTCTGCCGCGCGCGGTGGCCCGCGCTCGTG 1607

RESULT 12
 US-08-672-211-1/c
 Sequence 1, Application US/08672211
 Patent No. 5874273
 GENERAL INFORMATION:
 APPLICANT: Stephens, Len
 APPLICANT: Hawkins, Phillip T.
 TITLE OF INVENTION: G-BETA-GAMMA REGULATED
 PHOSPHATIDYLINOSITOL-3' KINASE
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 2730 Sand Hill Road
 CITY: Menlo Park
 STATE: California
 COUNTRY: USA
 ZIP: 94025
 COMPUTER READABLE FORM:

235 ggcctcggggactttctactcatctctcctggcaatctg - gaggcacaggcagggcaggt 293 QY
267 ggcctcggggactttctactcatctctcctggcaatctg - gaggcacaggcagggcaggt 325 QY
270 GCGGAGGGGGCGGTGACGTGGCGCAGGCGCGGGATTTCGCCGGCGGTGGGTAGCGGGT 211 Db
294 ggcgcgctgtcgtctccctcggggccgaaggagtgcacacacagctctctccggcgg 352 QY
316 ggcgcgctgtcgtctccctcggggccgaaggagtgcacacacagctctctccggcgg 384 QY
210 GCGGTGGCGGACAGCGGGGGGACGTGACAGGGGCGCAGAGCGGGGCGACGCGGGCGG 152 Db

RESULT 15

```

US-08-680-506-8/c
? Sequence 8, Application US/08680506C
? Patent No. 6008013
? GENERAL INFORMATION:
? APPLICANT: Reynolds, Paul R.
? TITLE OF INVENTION: CHONDROCYTE PROTEINS
? FILE REFERENCE: 176/60091
? CURRENT APPLICATION NUMBER: US/08/680,506C
? CURRENT FILING DATE: 1996-07-08
? EARLIER APPLICATION NUMBER: 60/021,672
? EARLIER FILING DATE: 1996-07-05
? NUMBER OF SEQ ID NOS: 18
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 8
? LENGTH: 1347
? TYPE: DNA
? ORGANISM: Gallus gallus
US-08-680-506-8

```

Query Match 2.5%; Score 38.2; DB 5; Length 1347;
Best Local Similarity 50.2%; Pred. No. 2.8;
Matches 120; Conservative 0; Mismatches 118; Indels 1; Gaps 1;

QY	115	c c o c g g t g t c c g t g a g t c c c c e l t t g t c g y g g a c c t g t g c c c c c a c a g c g a g c t g	174
D b	390	c c c c a g a g t g t t c c a t e n t c g c g c g c c c c g c g c c a g c c c g c t c g a g a g c t c t c t c	331
QY	175	t t i s g a a g t g g c c a t t c t g g y g g c g c c c g g a g g c g g c c c c g g g g t c c c g c	234
D b	330	g t a c c a c a t o g g c c a g c g c g c a c c g c g c g c g c g a g c a g c g c g c g c g c g c g	271
QY	235	g g c t c t g g g a c t t t o t a c t c a c c t g g c a a t c t y - g a g g c a a a g a g c a g g c a g t	293
D b	270	c g g a g a g c g c c g t g c a g c t g c c c a g c g c g g a t t g c c c g c c g t g g g t a g c g g t	211
QY	294	g g c c g c g t g t g c t c c c c g g g c g a a g a g t g y c c a a g a g c a g c t c t t c c g g g c g	352
D b	210	s c c c t g c g c a c a c g c g c g a g t g t c a g c g g c g c g a g c g c g c g c a g c g c g c g	152

Search completed: December 7, 2000, 13:40:01
Job time: 27450 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 10:56:07 ; Search time 152.48 Seconds
(without alignments)
3811.321 Million cell updates/sec

Title: us-09-434-708-1

Perfect score: 1547

Sequence: 1 cggaggcccccgtgctct.....caaaaaaaaaaaaaaaaaa 1547

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 950044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36.*

```

1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	340	22.0	3090	20 X89079	Nucleotide sequenc
2	48	3.1	10732	21 A10594	Gene encoding a su
3	46.8	3.0	114955	20 X53491	Human adenosine A1
4	46.4	3.0	9960	19 V58939	Mycobacterium smeg
C 5	43	2.8	114955	20 X53491	Human adenosine A1
C 6	42.2	2.7	1693	21 Z56005	Contig 004 from co
C 7	41.6	2.7	1140	21 Z87290	S. venezuelae deso
C 8	41.6	2.7	3291	21 Z56004	Contig 003 from co
9	41.6	2.7	12441	21 Z87284	S. venezuelae deso
10	41.6	2.7	15613	21 Z87319	S. venezuelae deso
11	41.4	2.7	1582	20 X08910	Human FADD protein
12	41.4	2.7	1642	17 T39397	FADD (Fas-associat

13	41.4	2.7	2644	20 X81941	cDNA encoding a pr
C 14	41.2	2.7	1170	20 Z39042	Human fishboy G-pr
C 15	40.4	2.6	880	17 T38913	B2-bradykinin rece
16	40.4	2.6	6225	20 X55273	Human enzyme-relat
17	40.4	2.6	6225	21 A34721	Human adenosine re
18	40	2.6	1701	17 T30372	MORT-1 cDNA. Homo
19	40	2.6	1701	18 T61397	MORT-1 coding sequ
20	40	2.6	1701	21 Z44745	Human FADD DNA. H
21	40	2.6	53789	19 V21187	Amycolatopsis medi
22	39.8	2.6	627	20 V1930	MORT1 isoform MORT
C 23	39.4	2.5	1308	17 T31602	Cartilage-derived
24	39.2	2.5	2170	21 Z94002	Porcine betal-adre
25	39.2	2.5	15672	12 Q10613	Rianodin receptor
C 26	38.8	2.5	738	20 X86968	PHA acetoacetyl-Co
27	38.8	2.5	3632	19 V61586	Alpha-1A calcium c
C 28	38.8	2.5	6436	20 X36965	DNA containing PHA
C 29	38.8	2.5	77536	21 A14651	Nucleotide sequenc
30	38.4	2.5	1746	11 Q05558	Sequence encoding
31	38.4	2.5	2914	20 X37059	Human KCNQ3 cDNA.
C 32	38.4	2.5	4692	19 V04633	Porcine phospholip
C 33	38.4	2.5	4692	20 V74099	Porcine G-protein
C 34	38.4	2.5	4692	21 Z86812	Pig p101 regulator
C 35	38.2	2.5	954	19 V19862	Gallus domesticus
C 36	38.2	2.5	1347	19 V19864	Gallus domesticus
C 37	38.2	2.5	1387	19 V19863	Gallus domesticus
C 38	38.2	2.5	2233	19 V19861	Gallus domesticus
C 39	38.2	2.5	5027	19 V19860	Sequence encoding
C 40	38	2.5	1999	11 Q04784	Mouse N-acetylgluc
C 41	38	2.5	2150	20 X87820	Tn5099-10 transpon
42	38	2.5	3318	15 Q68194	Sequence comprisin
C 43	38	2.5	5392	15 Q84201	MH mutant porcine
C 44	38	2.5	15377	13 Q25975	Platenolide synth
45	38	2.5	44377	18 T78508	

ALIGNMENTS

RESULT 1

X89079
ID X89079 standard; DNA; 3090 BP.

XX AC X89079;

XX DT 14-SEP-1999 (first entry)

XX DE Nucleotide sequence of human Cbl.

XX KW LAT; tyrosine kinase; linker for activation of T cell; TCR; human;
KW T-cell receptor; TCR signalling pathway; neoplasia; inflammation;
KW hypersensitivity; allergy; microbial infection; genetic disease;
KW autoimmune disease; graft rejection; modulator; Cbl; ss.

OS Homo sapiens.

XX PN WO9932627-A2.

XX PD 01-JUL-1999.

XX PF 23-DEC-1998; 98WO-US27400.

XX PR 23-DEC-1997; 97US-0068690.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Samelson LE, Zhang W;

XX DR WPI; 1999-418926/35.

XX DR P-PSDB; Y27126.

XX PI Linker for activation of T cell protein used to, e.g. screen for
PI modulators of T cell signalling

XX XX

[illegible][illegible]

xx This is contig 304 from the recombinant cosmid PKOS023-27 DNA sequence
cc (see 256001) which contains a *Streptomyces venezuelae* DNA insert. The
cc cosmid contains open reading frames which encode the various modules of
cc the narbonolide polyketide synthase (PKS). The invention relates to
cc the narbonolide polyketide synthase (PKS). The invention relates to
cc recombinant DNA containing a coding sequence for a narbonolide PKS.
cc Polyketides are compounds synthesised from 2-carbon units through a
cc series of condensations and subsequent modifications. Modular PKSs are
cc responsible for the production of many antibiotics including piromycin.
cc The narbonolide PKS consists of a loading module, six extender modules,
cc

CC and two thioester domains. Four proteins make up the narbonolide PKS
 CC (PICAI, PICAI, PICAI, and PICAI). PICAI includes the loading module
 CC and extender modules 1 and 2, PICAI includes extender modules 3 and 4,
 CC PICAI includes extender module 5 and PICAI includes extender module 6
 CC and a type II thioesterase domain. The second type II thioesterase domain
 CC is found on the PICB protein. The nucleotide sequences encoding all of
 CC these proteins can be isolated in recombinant form from the recombinant
 CC cosmid pKOS23-27. Narbonolide is desaminylated in S. venezuelae to
 CC yield narbornolide, and the desaminyl transferase enzyme is required for this
 CC conversion, and the desaminyl transferase genes are also found in
 CC cosmid pKOS23-27. The recombinant DNA of the invention is used to
 CC express, in transformed cells, narbonolide (or its derivatives) or other
 CC ketolides (particularly hybrids), which may then be converted (e.g. by
 CC other enzymes recombinantly expressed in the same hosts) to polyketide
 CC antibiotics or their intermediates. The antibiotics are useful in human
 CC or veterinary medicine.
 XX
 SQ Sequence 1693 BP; 237 A; 566 C; 633 G; 257 T; 0 other;

Query Match 2.7%; Score 42.2; DB 21; Length 1693;
 Best Local Similarity 48.9%; Pred. No. 1.7; Indels 0; Gaps 0;
 Matches 113; Conservative 0; Mismatches 118;
 QY 809 gctgtactcgcctggggcagtgccatcggtatgtgactcagatggcagcagctctgc 868
 DB 971 GCTGCGAGTCCAGTCCGCTAGCGATCGGCAAGCCGAGCCGCTCGCCCTTCGTCG 912
 QY 869 agaccatctcgcacaaacccctctccagtgctctcagtgaggagacagagcgct 928
 DB 911 AGACCTTCGCGACCCCAACATGATGAGACCGGAGAGATCGAGAACGCCATCGCGAGGCT 852
 QY 929 tctacctctaccagatggaagaccacacacacagctgactgagctcgccagggcag 988
 DB 851 TCGACCTCGCGCGCGGGGATCATCCGAGACTCGACCTGCTCGCCGATCTACTCC 792
 QY 989 aaccacagcagcgatccatcgctgtagagagcagctgagctctactctggg 1039
 DB 791 AGACCCGCGCTACGCCCACTTCGCGCGGAGCTCGCGGACTTCACCTGGG 741

RESULT 7
 287290/c
 ID 287290 standard; DNA; 1140 BP.
 XX AC 287290;
 XX DT 05-JUN-2000 (first entry)
 XX DE S. venezuelae desosamine biosynthetic gene desV, SEQ ID NO:15.
 XX KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KW neomethymycin; narbornolide; polyhydroxyalkanoate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolemia; crop protection agent; ds.
 XX OS Streptomyces venezuelae ATCC15439.
 XX FH Key Location/Qualifiers
 FT CDS 1..1140
 FT /*tag= a
 FT /product= "Desv"
 XX WO200000620-A2.
 XX PD 06-JAN-2000.
 XX PF 25-JUN-1999; 99WO-US14398.
 XX PR 26-JUN-1998; 98WS-0105537.
 XX PA (MINS) UNIV MINNESOTA.

XX Sherman DE, Liu H, Xue Y, Zhao L;
 XX WPI: 2000-160679/14.
 DR P-PSDB; Y77185.
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
 PT synthesis of methymycin and pikromycin -
 XX Claim 3; Page 362; 438pp; English.
 XX The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the eryl gene cluster of Saccharopolyspora erythraea or
 CC Streptomyces antibioticus. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
 CC pikromycin, neomethymycin, narbornolide or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the production
 CC of biologically active macrolides. The macrolide biosynthetic proteins
 CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
 CC narbornolide. The alternative termination of polyketide synthase may be
 CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
 CC monomers. The compounds produced by the recombinant host cells are useful
 CC as biopolymers, e.g., in packaging or biomedical applications, to
 CC engineer PHA monomer synthases or to prepare biologically active agents,
 CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
 CC chronic obstructive pulmonary disease as well as other diseases involving
 CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
 CC antibiotics which are active against a variety of organisms, e.g.,
 CC bacteria, including multi-drug resistant pneumococci and other
 CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
 CC protection agents (e.g., fungicides or insecticides) via expression of
 CC polyketides in plants. Sequences 287286-287294 represent desosamine
 CC biosynthetic genes from Streptomyces venezuelae ATCC 15439, which encode
 CC proteins Y77181-Y77189.
 XX
 SQ Sequence 1140 BP; 163 A; 472 C; 373 G; 132 T; 0 other;

Query Match 2.7%; Score 41.6; DB 21; Length 1140;
 Best Local Similarity 46.3%; Pred. No. 2.1;
 Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
 QY 85 ctgcagcgctcagaagacgaatcgtcagccccccgctcgtcgtgagtcctcccttcgctg 144
 DB 1125 CCGCTCGGCCCATTCGGCGACGGCTCGATCACCCAGCGGCTCGGGGCTCCAGGTG 1066
 QY 145 cgggacatcgtcgcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagc 204
 DB 1065 CGGCGCATCGGACGGCTGAGGACCTCGCGCGAAGCTCTCGCGCGCGGAGCGAGCC 1006
 QY 205 gcggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 264
 DB 1005 TTCCGCGGTTCCTTCGCCCGCTAGGCGGCGAGAGGTGCACGGGTACCGGGTAGTGC 946
 QY 265 gccaatctggagcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 324
 DB 945 GAGGGTGTGCATGCCGCGGCGCTCGAGGTGGTGCAGAGCTCGTCGCGGGGCTCGGTG 786
 QY 325 agtgccaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 380
 DB 885 CACCGTGAGAGGTGCCAGACCGGGTGGTGTGCGGCGCGGTACCGGCGAGCGCGCA 830

RESULT 8
 ID 256004/c
 XX 256004 standard; DNA; 3291 BP.
 AC 256004;
 XX

DT 23-MAR-2000 (first entry)
XX
DE Contig 003 from cosmid PKOS023-27 from Streptomyces venezuelae.
XX
KW Narbonolide polyketide synthase; PKS; cosmid PKOS023-27; contig 003;
KW ketolide; dNDP glucose synthase; dehydratase; picromycin; narbonolide;
KW antibiotic production; ds.
XX
OS Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT CDS 104..982
FT CDS /*tag= a
FT /*product= dNDP_glucose_synthase
FT /*note= "glucose-1-phosphate thymidyl transferase"
FT CDS 1114..2127
FT /*tag= b
FT /*product= dNDP_glucose4_6_dehydratase
FT CDS 2124..3263
FT /*tag= c
FT /*product= PICCI
FT /*transl_except= (Pos:2874..2876, aa:Xaa)
FT /*note= "Xaa = Unknown"
XX
PN WO9961599-A2.
XX
PD 02-DEC-1999.
XX
XX 27-MAY-1999; 99WO-US11814.
XX
XX 28-MAY-1998; 98US-0087080.
XX 28-AUG-1998; 98US-0141908.
XX 22-SEP-1998; 98US-0100880.
XX 08-FEB-1999; 98US-0119139.
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
PI WPI; 2000-072618/06.
DR P-PSDB; Y67206, Y67214, Y67215.
DR
XX
PT New recombinant DNA encoding a domain of narbonolide polyketide
PT synthase, for production of ketolide antibiotics -
XX
XX Disclosure; Page 37-38; 98pp; English.

XX
XX This is contig 003 from the recombinant cosmid PKOS023-27 DNA sequence
XX (see Z56001) which contains a Streptomyces venezuelae DNA insert. The
XX cosmid contains open reading frames which encode the various modules of
XX the narbonolide polyketide synthase (PKS). The invention relates to
XX recombinant DNA containing a coding sequence for a narbonolide PKS.
XX Polyketides are compounds synthesised from 2-carbon units through a
XX series of condensations and subsequent modifications. Modular PKSs are
XX responsible for the production of many antibiotics including picromycin.
XX The narbonolide PKS consists of a loading module, six extender modules,
XX and two thioester domains. Four proteins make up the narbonolide PKS
XX (PICAI, PICAI, PICAI and PICAI). PICAI includes the loading module
XX and extender modules 1 and 2, PICAI includes extender modules 3 and 4,
XX PICAI includes extender module 5 and PICAI includes extender module 6
XX and a type II thioesterase domain. The second type II thioesterase domain
XX is found on the PICB protein. The nucleotide sequences encoding all of
XX these proteins can be isolated in recombinant form from the recombinant
XX cosmid PKOS023-27. Narbonolide is desaminylated in S. venezuelae to
XX yield narbonolide. The desaminyl transferase enzyme is required for this
XX conversion, and the desaminase biosynthetic genes are also found in
XX cosmid PKOS023-27. The recombinant DNA of the invention is used to
XX express, in transformed cells, narbonolide (or its derivatives) or other
XX ketolides (particularly hybrids), which may then be converted (e.g. by
XX other enzymes recombinantly expressed in the same hosts) to polyketide
XX antibiotics or their intermediates. The antibiotics are useful in human
XX or veterinary medicine.

SQ Sequence 3291 BP; 521 A; 1289 C; 1044 G; 436 T; 1 other;
Query Match 2.7%; Score 41.6; DB 21; Length 3291;
Best Local Similarity 46.3%; Pred. No. 2.7;
Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
QY 85 ctgcagcgcctagaagcgaatgcgtgcagcccccccggtgtccgtgagtcocccctcgctg 144
DB 3247 CCGCTCGGCCCAATTCGGCCACGGCGTCGATCACCGCAGGCGCTCGCGGCTCCAGGTG 3188
QY 145 cgggacactgtgccccgcacagcgacgtcttcgagagtggtgcccatttcgagcgggcg 204
DB 3187 CGGCCCATGCGCAGGCTGAGGACCTGCCGCGGCGAGAGTGCAGGGTACCGGGTAGTCGT 3068
QY 205 gcggcgaggcgccccgggggtcccgcggtctctgggacttttactactatcactg 264
DB 3127 TTCGGCGGCTGCTCGCCGCGTAGCGGCGGCGAGAGTGCAGGGTACCGGGTAGTCGT 3068
QY 265 gccaatctgagggccaaagacagcagcagtggtggcgcgctgctcccgggcggaagg 324
DB 3067 GAGGTGTCGATGCCGCGGCGGCGTGCAGGTGCTCGCGAGCTGTCGCGGCGCTCGGTGCG 3008
QY 325 agtgcacacgacgagctcttcggcggtcccgagactcagcgagcagctggccaa 380
DB 3007 CACGTGAGAGAGTGCAGACCGCGGTGCGTTCGGCGCGGTACCGCGCAGGCCGA 2952
RESULT 9
Z87284
ID Z87284 standard; DNA; 12441 BP.
XX
AC Z87284;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae desosamine biosynthetic gene cluster pikB, SEQ ID NO:3.
KW Desosamine biosynthesis; macrolide; polyketide; picromycin; pikromycin;
KW neomethymycin; narbonolide; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-150679/14.
DR P-PSDB; Y77179.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT synthesis of methymycin and pikromycin -
XX
PS Claim 2; Page 281-287; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
XX comprising a desosamine biosynthetic gene cluster, a fragment or its
XX biologically active variant, where the nucleic acid sequence is not
XX derived from the erylC gene cluster of Saccharopolyspora erythraea or
XX Streptomyces antibioticus. The invention also relates to a macrolide
XX biosynthetic gene cluster, or fragments thereof. The macrolide
XX biosynthetic gene cluster encodes proteins which synthesize methymycin,
XX

CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the production
 CC of biologically active macrolides. The macrolide biosynthetic proteins
 CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
 CC narbomycin. The alternative termination of polyketide synthesis may be
 CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
 CC monomers. The compounds produced by the recombinant host cells are useful
 CC as biopolymers, e.g., in packaging or biomedical applications, to
 CC engineer PHA monomer synthases or to prepare biologically active agents,
 CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
 CC chronic obstructive pulmonary disease as well as other diseases involving
 CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
 CC antibiotics which are active against a variety of organisms, e.g.,
 CC bacteria, including multi-drug resistant pneumococci and other
 CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
 CC protection agents (e.g., fungicides or insecticides) via expression of
 CC polyketides in plants. The present sequence represents the desosamine
 CC biosynthetic gene cluster from Streptomyces venezuelae ATCC 15439.
 XX
 SQ Sequence 12441 BP; 1704 A; 4294 C; 4686 G; 1757 T; 0 other;

Query Match 2.7%; Score 41.6; DB 21; Length 12441;
 Best Local Similarity 46.3%; Pred. No. 3.5;
 Matches 137; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 85 ctgcagcgctagaagagcaatcgctgaccccccggctgctgagtcaccccttcgtg 144
 Db 6158 cctgctggccatctgcgcagcgctgcatccacccagcgctgcggcgctccaggcg 6217
 QY 145 cggagacctgctccgcgcagcgagctgcttcgagagtggtgcccattctcggggcg 204
 Db 6218 cggcgcatcgagcgctgaggaacctgcgcgcagagctcgcgcgggagcgagcc 6277
 QY 205 gccggcgagggcgcccggggtcccgcggtctcgggactttctactatctacctg 264
 Db 6278 ttccggcggtgctcgcccgctagcgcgcgagagtgacgggtaccgggtagtgctg 6337
 QY 265 gccaatctgagcccaagacagcgagtggtggtccgctgctcccgggcgagcg 324
 Db 6338 gagggtgctgatcccgcgggcgctcgagggggtcgcgagctcgcgcgctcggcg 6397
 QY 325 agtgcacacgacgagctctcccgcggggtccagactcagactcagcgagcgagtgccaa 380
 Db 6398 cagcgtgagaggtgcagacggggtcggtgctcgggcggtcaccggcgagggcgga 6453

RESULT 10
 Z87319
 ID Z87319 standard; DNA; 13613 BP.
 AC Z87319;
 XX
 XX Z87319;
 XX
 XX 05-JUN-2000 (first entry)
 XX
 DE S. venezuelae desosamine biosynthetic gene cluster pikB.
 XX
 KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolaemia; crop protection agent; ds.
 XX
 OS Streptomyces venezuelae ATCC15439.
 XX
 XX Key Location/Qualifiers
 FH 1..809
 FT CDS
 FT /tag= a
 FT /partial
 FT /product= "pikB gene cluster protein #1 (Y77204)"
 FT /note= "No initiation codon given in the specification"
 FT CDS . 806..2014

FT
 FT
 FT CDS
 FT /tag= b
 FT /product= "pikB gene cluster protein #2 (Y80998)"
 FT 2162..6741
 FT /tag= c
 FT /product= "pikB gene cluster protein #3 (Y77205)"
 FT 6834..7402
 FT /tag= d
 FT /product= "pikB gene cluster protein #4 (Y77206)"
 FT /transl_except= (pos:6837..6841, aa:Glu)
 FT 7492..8205
 FT /tag= e
 FT /product= "pikB gene cluster protein #5 (Y77207)"
 FT /complement (7942..8205)
 FT /partial
 FT /tag= f
 FT /product= "pikB gene cluster protein #6 (Y77208)"
 FT /note= "No termination codon given in the specification"
 FT /transl_except= (pos:8270..8272, aa:Val)
 FT /transl_except= (pos:8273..8275, aa:Thr)
 FT /transl_except= (pos:8276..8278, aa:Gly)
 FT /complement (10126..11139)
 FT /tag= g
 FT /product= "pikB gene cluster protein #7 (Y80999)"
 FT /complement (11271..12149)
 FT /tag= h
 FT /product= "pikB gene cluster protein #8 (Y77209)"
 FT /complement (12342..13799)
 FT /tag= i
 FT /product= "pikB gene cluster protein #9 (Y77210)"
 FT /complement (13706..15043)
 FT /tag= j
 FT /product= "pikB gene cluster protein #10 (Y77211)"
 FT 15404..15574
 FT /tag= k
 FT /product= "pikB gene cluster protein #11 (Y77212)"
 XX
 XX WO200000520-A2.
 XX
 XX 06-JAN-2000.
 XX
 XX 25-JUN-1999; 99WO-US14398.
 XX
 XX 26-JUN-1998; 98US-0105537.
 XX
 XX (MINU) UNIV MINNESOTA.
 XX
 XX Sherman DH, Liu H, Xue Y, Zhao L;
 XX
 XX WPI: 2000-150679/14.
 XX
 XX P-PSDB: Y77204, Y77205, Y77206, Y77207, Y77208, Y77209,
 XX Y77210, Y77211, Y77212, Y80998, Y80999.
 XX
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
 XX synthesis of methymycin and pikromycin -
 XX
 XX Disclosure; Figure 32; 438pp; English.
 XX
 XX The invention relates to an isolated and purified nucleic acid segment
 XX comprising a desosamine biosynthetic gene cluster, a fragment or its
 XX biologically active variant, where the nucleic acid sequence is not
 XX derived from the ercC gene cluster of Saccharopolyspora erythraea, or
 XX Streptomyces antibioticus. The invention also relates to a macrolide
 XX biosynthetic gene cluster, or fragments thereof. The macrolide
 XX biosynthetic gene cluster encodes proteins which synthesize methymycin,
 XX pikromycin, neomethymycin, narbomycin or a combination of these
 XX compounds. Recombinant or augmented cells comprising the desosamine
 XX and/or macrolide biosynthetic gene clusters are useful for the production
 XX of biologically active macrolides. The macrolide biosynthetic proteins
 XX are useful for synthesis of methymycin, pikromycin, neomethymycin and
 XX narbomycin. The alternative termination of polyketide synthesis may be
 XX useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
 XX monomers. The compounds produced by the recombinant host cells are useful
 XX as biopolymers, e.g., in packaging or biomedical applications, to

engineer PHA monomer synthases or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polypeptides in plants. The present sequence represents the desosamine biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC 15439, as given in figure 32.

Query Match	2.78;	Score 41.6;	DB 21;	Length 13613;
Best local Similarity	40.3%;	Pred. NO. 3.6;		
Matches 137;	Conservative 0;	Mismatches 159;	Indels 0;	Gaps 0;
QY	85	ctgcagcgcttagaagacgaatcgctcgaccccccgctgtccgtgagtcgcccttcgctg	144	
Db	6845	ccgctcggcccatctcgcgacggcgctcgataccccgagcgctcgggcggtccacggtg	6904	
QY	145	cgggactgtgcgcccgacacagcgacactgtctcgagaggtggccattctcggggcg	204	
Db	6905	cgggcccgtatcgcgaggtctgaggaactgccgcgcggaagctctcgcccgcgagcgagtc	6964	
QY	205	gcgcgcgagcgcgcccccggggtcccgcgcgctctgggacatttactatctacctg	264	
Db	6965	ttcgcgggtgcctcgcccgcgtaggcggcgagaggtgcacgggtacccgggtagtgcg	7024	
QY	265	gccaatctgaggggccaaagacagcaggtgagcgcgctgctgcccccggggccgaagg	324	
Db	7025	gagggtgtcgatgcgcggggcgctcgaggttgctgcgcagctctctcgcgctctcggtcg	7084	
QY	325	agtgcacacgcagagctcttccggcgggggtccagactcaggcagcagctggccaa	380	
Db	7085	cacggtgaagaggtgcccagacagcggttcggtgttcgggcggtcacggcgagcgga	7140	

RESULT 11
X08910
ID X08910 standard; cDNA; 1582 BP.

XX		
AC	X08910;	
XX		
XX		
DT	27-APR-1999	(first entry)
XX		
DE	Human FADD protein coding sequence.	

KW FIP; FADD interacting protein; FADD; Fas-associated protein with a
 novel death domain; cell death; apoptosis; Alzheimer's disease;
 KW Acquired Immune Deficiency Syndrome; AIDS; muscular dystrophy;
 KW amyotrophic lateral sclerosis; virus; bacteria; fungus; mycoplasma;
 KW protozoa; neoplasia; dysplasia; hyperplasia; ds.
 XX
 OS Homo sapiens.
 OS

XX	Key	Location/Qualifiers
FH	CDS	70..696
FT		

FT /label= Human_FADD_protein

PN WO9900499-A1.

07-JAN-1999.

26-JUN-1998:

03-JUN-1998:

PR 20-JUN-1997;
XX

PA (CHIR) CHIR

XX	Chen TT, Williams LT;
PI	WPI; 1999-095745/08.
XX	DR P-PSDB; W96154.
XX	XX
PT	New FADD (Fas-associated protein with a novel death
PT	domain)-interacting Protein - useful for inducing or preventing
PT	apoptosis in a cell, to aid in controlling apoptosis-related
PT	diseases
XX	XX
PS	Dislosure; Page 45-46; 58pp; English.

An epitope of human FADD (Fas-associated protein with a novel death domain)-interacting protein (FIP protein) comprising amino acids 348-727 of the protein described in W96153, can be used to induce or prevent apoptosis in a cell. Specifically, decreasing the levels of FIP348-727 prevents apoptosis. This is useful in cells which are dying prematurely, eg: Alzheimer's disease, Acquired Immune Deficiency Syndrome (AIDS), muscular dystrophy, anorectic lateral sclerosis (and other muscle wasting diseases), autoimmune diseases, and diseases where cells are infected with a pathogen (virus, bacteria, fungus, mycoplasma or protozoa). Increasing the levels of FIP 348-727 induces apoptosis which is useful in cells suffering from neoplasias, dysplasias, hyperplasias, or their symptoms. Purified and isolated FIP subgenomic polynucleotides are useful as primers to obtain more copies of the nucleotides, and as probes that identify wild-type or mutant coding sequences. They are also useful for expressing FIP mRNA, proteins or fusion proteins, and in the generation of FIP antisense oligonucleotides and ribozymes. They are also useful in expression constructs and gene delivery vehicles (optionally in combination with a condensing agent) that deliver FIP mRNA or oligonucleotides, FIP-protein (including variants), FIP-specific ribozymes or single-chain antibodies into eukaryotic cells. This is the human FADD protein gene. Human FIP protein binds to amino acids 1-110 of the human FADD protein given in W96154.

```

Query Match      2.7%; Score 41.4; DB 20; Length 1582;
Best Local Similarity 48.9%; Pred. No. 2.5;
Matches 111; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```

33	gtggggcgacagtggaagagccgcgcgccttgccggcgagtgctgcagcg	92
QY		
156	ggcgtgggcgaagcgcaagtggagcgtgcagagcgccctagaccctcttcacgt	215
Db		
93	cctagaagacaatgcgtcaccgccgtgtccgtgagtcgcccttcgcgcgggaact	152
QY		
216	gctggagcagaaacacctggagccggggacacccgagctcctgcgcgagctgcgcctc	275
Db		
153	gtgcccgcgcacagcgagctgtcttcagaggtgggccattctcggcggggcgccggcg	212
QY		
276	cctgcggcgccacgaacctgctgcggcgctcgcagacttcgagggcgggcgccggcg	335
Db		
213	aggcggcccggggggtcccgcgcgctctgggggactttactcatct	259
QY		
336	agccgcgctgggaagaagaacacctgtctgcagcatttaacgtcatat	382
Db		

RESULT 12

T39397
ID T39397 standard: cDNA: 1642 BP.

XX
AC T39397:XX
22-FEB-XX
DE
FADD (FXX
 11/23/54

2

KW apoptosis; Fas receptor; death domain; gene therapy; antibody;
 KW immunosay; drug screening; diagnostic; AIDS; antiinflammatory;
 KW antitumor; cerebroprotective; neuroprotective; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 5'UTR 1..129
 FT misc_feature 4..6
 FT /tag= a
 FT /tag= b
 FT /note= "In-frame stop codon"
 FT CDS 130..756
 FT /tag= c
 FT /product= FADD protein
 FT misc_feature 133..501
 FT /tag= d
 FT /note= "Encodes N-terminal half, inducing apoptosis
 but not binding Fas receptor"
 FT misc_feature 198
 FT /tag= e
 FT /note= "Clone 15 start point"
 FT misc_feature 249
 FT /tag= f
 FT /note= "Clone 8 start point"
 FT misc_feature 460..660
 FT /tag= g
 FT /note= "Region encoding death domain"
 FT 3'UTR 757..1642
 FT /tag= h
 FT polyA_signal 1636..1641
 FT /tag= i
 XX
 PN W09631603-A2.
 XX
 PD 10-OCT-1996.
 XX
 XX 28-FEB-1996; 96WO-0502857.
 XX 18-MAY-1995; 95US-0443982.
 PR 03-APR-1995; 95US-0416379.
 XX
 PA (UNMI) UNIV MICHIGAN.
 XX
 XX Dixit VM, O'Rourke K;
 PI WPI; 1996-465026/46.
 DR P-PSDB; W03653.
 XX
 XX FADD protein that binds to cytoplasmic region of Fas receptor - for
 PT identifying inhibitors of Fas-associated apoptosis useful for
 PT treating e.g. AIDS, leukemia, stroke, etc
 XX
 PS Example 1; Fig 2A-B; 96pp; English.
 XX
 CC The sequence encodes FADD (Fas-associating protein with novel death
 CC domain), which binds the cytoplasmic region of a Fas receptor, and
 CC modulates apoptosis induced by activation of the receptor by ligand
 CC binding. The cDNA has been isolated using a yeast two-hybrid system
 CC which screens for proteins interacting with the Fas cytoplasmic
 CC domain. A GAL4 DNA-binding domain has been fused to the human Fas
 CC antigen cytoplasmic tail to form a bait plasmid, which is used with
 CC a prey plasmid, containing a human B-lymphocyte cDNA library fused to
 CC the GAL4-activation domain, to co-transform yeast cells.
 CC Overlapping clones 8 and 15 have been isolated, and the full-length
 CC cDNA has been isolated from a human umbilical vein endothelial cell
 CC (HUVEC) library using clone 15 as a probe. The cDNA has an
 CC in-frame stop codon 130 bp upstream of the Initiator Met. The
 CC encoded protein contains a death domain, with interacts with the
 CC death domain of Fas. The DNA may be used in gene therapy, and the
 CC protein or a corresponding antibody may be used to screen for
 CC agents modulating FADD pathway cellular functions and Fas-associated
 CC apoptosis, for use in therapy of e.g. AIDS, inflammation,

CC Leukaemia, myocardial infarction, degenerative disease, etc.
 XX
 SQ Sequence 1642 BP; 354 A; 448 C; 508 G; 332 T; 0 other;
 XX
 Query Match 2.7%; Score 41.4; DB 17; Length 1642;
 Best Local Similarity 48.9%; Pred. No. 2.6;
 Matches 111; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
 QY 33 gtggggggaagtagtgggaagggccgcgcctggggcggcagtcaggtgctgcagcg 92
 Db 216 ggcgtggggaagcgaagctggagcgtgcagagcgtcagacacctctctccatgct 275
 QY 93 cctagaagagcgaatgcgtcgaccccggtgtccgtgagtcctcctcgtgcgggacct 152
 Db 276 gctggagcagracgacctggagccggccacacgagctcctgcgagctgcgctc 335
 QY 153 gctgcccgcacagcgcgtcgttcgagagtggtggccattctctggggcgccggcg 212
 Db 336 cctggcggaagcagacctgctgcggcgtcgcagcacttcgagcgggcgccggcg 395
 QY 213 agcgcgcccggggtcccgcggtctgtgggactttctactatct 259
 Db 396 ggcgcgcctggggaagacctgtgtgcagcatttaacgtcatat 442
 RESULT 13
 X81941
 ID X81941 standard; cDNA; 2644 BP.
 XX
 AC X81941;
 XX
 DT 09-SEP-1999 (first entry)
 XX
 DE cDNA encoding a protein involved in eliciting a signal in HH-PTC.
 XX
 KW Human homologue; Drosophila fused gene; intracellular signal;
 KW human hedgehog-patched protein; HH-PTC; RNA transcription;
 KW RNA translation; cancer diagnostic; prophylaxis; karyotyping analysis;
 KW cancer; embryonic repair; tissue repair; wound healing;
 KW neurodegenerative disease; testicular function; ss.
 XX
 OS Homo sapiens.
 XX
 PW W09932609-A1.
 XX
 PD 01-JUL-1999.
 XX
 PF 18-DEC-1998; 98WO-SE02384.
 XX
 PR 26-JUN-1998; 98SE-0002292.
 PR 19-DEC-1997; 97SE-0004788.
 XX
 PA (KARO-) KAROLINSKA INNOVATIONS AB.
 XX
 PI Toftgard R, Zaphiropoulos PG;
 XX
 DR WPI; 1999-418918/35.
 DR P-PSDB; Y23755.
 XX
 PT Human homologue of the Drosophila fused gene
 XX
 PS Claim 11; Page 74-77; 81pp; English.
 XX
 CC The present sequence encodes a human homologue to molecules associated
 CC with the Drosophila fused gene. The protein is involved in eliciting an
 CC intracellular signal in the human hedgehog-patched (HH-PTC). The
 CC protein, its antibodies and pharmaceutical compositions comprising them
 CC can be used as medicaments. The presence of a fused gene, cDNA, mRNA,
 CC protein or subsequences of these in a biological sample are useful,
 CC e.g. as a marker to assess in vivo and/or in situ RNA transcription
 CC and/or translation, in cancer diagnostics, in prophylaxis, etc. The
 CC polynucleotide sequence can also be used to derive probes for the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

395 CACCGTAGTACACCGCTTCAGCCCGCTGGCCACGGCCTGCCGG 342

myocardial infarction; coronary heart disease; cardiomyopathy; KW

KW hypertension; genetically related vascular disease; sterility;
KW infertility; atopia; asthma; allergy; altered bronchial reactivity;
KW acute; chronic; inflammation; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
XX repeat_region 117..841
XX FT /*tag= a
XX FT /note= "Contains 48 tandem repeats with the
XX FT allele 117..841
XX FT /*tag= b
XX FT /label= BE3-48R
XX FT /note= "alternative allele BE3-35R contains 35
XX FT tandem repeats with the consensus sequence
XX FT TGA(A)GGGCTAGAACC"
XX FT repeat_unit 117..131
XX FT /*tag= c
XX FT /rpt_type= TANDEM

XX DE19518931-Cl.

XX PN

XX XX

XX PD

XX 19-SEP-1996.

XX 23-MAY-1995; 95DE-1018931.

XX 23-MAY-1995; 95DE-1018931.

XX (ROSC/) ROSCHER A.

XX Braun A, Kammerer S;

XX WPI; 1996-426182/43.

XX New polymorphic regions in the B2-bradykinin receptor gene - for
XX diagnosis and treatment of bradykinin related diseases, e.g.
XX myocardial infarction and allergy, and for drug assessment

XX Claim 4: Fig 3: 13pp; German.

XX The present sequence represents allele BE3-R48 of the tandem-repeat
XX polymorphic region of exon 3 of the B2-bradykinin receptor gene.
XX Two alleles have been identified; allele BE3-48R contains 48
XX tandem copies of repeat units with the consensus sequence
XX TGA(A)GGGCTAGAACC, while allele BE3-35R contains 35 copies.
XX The polymorphic sequences are useful for diagnosis (i.e. relating
XX frequency of a polymorphism to disease occurrence) and/or treatment
XX of myocardial infarction; coronary heart disease; cardiomyopathy;
XX hypertension; genetically related vascular disease; sterility;
XX infertility; atopia; asthma; allergy; altered bronchial reactivity,
XX and acute or chronic inflammation. They can also be used to test
XX potential drugs against such diseases and pharmaceuticals that contain
XX angiotensin-converting enzyme inhibitors, kinin-receptor (ant)agonists
XX and bradykinin (ant)agonists.

XX Sequence 880 BP; 277 A; 170 C; 299 G; 134 T; 0 other;

Query Match 2.6%; Score 40.4; DB 17; Length 880;
Best Local Similarity 53.1%; Pred. No. 3.7;
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1023 gctgcagctctactggccatggacccacatttgagctctgacagatctgtgtgagag 1082
DB 401 GTTCTAGCTCTCCAGGTTCTAGCCCTCCAGGTTCTAGCTCTCCAGGTTCTAGCCCTCCAG 342
QY 1083 caacaggatgtgaagattgagccgtggcgacctgctctgagctgctgctgctgc 1142
DB 341 GTTCTAGCTCTCCAGGTTCTAGCCCTCCAGGTTCTAGCCCTCCAGGTTCTAGCTCTCCAG 282
QY 1133 ctgggagcactcggacagcagcactgccccttctgcgctg 1184

Db 281 ATTCCAGCCCTCCAGGTTCTAGCTCTCCAGGTTCTAGCTCTCCAG 240

Search completed: December 7, 2000, 18:06:44
Job time: 25837 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 12:41:38 ; Search time 1490.04 Seconds
(without alignments)
5900.480 Million cell updates/sec

Title: us-09-434-708-3
Perfect score: 1422
Sequence: 1 atggtcttgccgtggccccc.....ccagagaccctgcccggcc 1422

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: gb_est1.*
 - 2: gb_est2.*
 - 3: gb_est3.*
 - 4: gb_est4.*
 - 5: gb_est5.*
 - 6: gb_est6.*
 - 7: gb_est7.*
 - 8: gb_est8.*
 - 9: gb_est9.*
 - 10: gb_est10.*
 - 11: gb_est11.*
 - 12: gb_est12.*
 - 13: gb_est13.*
 - 14: gb_est14.*
 - 15: gb_est15.*
 - 16: gb_est16.*
 - 17: gb_est17.*
 - 18: gb_est18.*
 - 19: gb_est19.*
 - 20: gb_est20.*
 - 21: gb_est21.*
 - 22: gb_est22.*
 - 23: gb_est23.*
 - 24: gb_est24.*
 - 25: gb_est25.*
 - 26: gb_est26.*
 - 27: gb_est27.*
 - 28: gb_est28.*
 - 29: gb_est29.*
 - 30: gb_est30.*
 - 31: gb_est31.*
 - 32: gb_est32.*
 - 33: gb_est33.*
 - 34: gb_est34.*
 - 35: gb_est35.*
 - 36: gb_est36.*
 - 37: gb_est37.*
 - 38: gb_est38.*
 - 39: gb_est39.*
 - 40: gb_est40.*
 - 41: em_estba.*
 - 42: em_estfun.*
 - 43: em_esthum1.*

- 44: em_esthum2.*
- 45: em_esthum3.*
- 46: em_esthum4.*
- 47: em_esthum5.*
- 48: em_esthum6.*
- 49: em_esthum7.*
- 50: em_esthum8.*
- 51: em_esthum9.*
- 52: em_esthum10.*
- 53: em_esthum11.*
- 54: em_esthum12.*
- 55: em_esthum13.*
- 56: em_esthum14.*
- 57: em_esthum15.*
- 58: em_esthum16.*
- 59: em_esthum17.*
- 60: em_esthum18.*
- 61: em_esthum19.*
- 62: em_esthum20.*
- 63: em_estin1.*
- 64: em_estin2.*
- 65: em_estin3.*
- 66: em_estin4.*
- 67: em_estom.*
- 68: em_estov1.*
- 69: em_estov2.*
- 70: em_estpl1.*
- 71: em_estpl2.*
- 72: em_estpl3.*
- 73: em_estpl4.*
- 74: em_estpl5.*
- 75: em_estro1.*
- 76: em_estro2.*
- 77: em_estro3.*
- 78: em_estro4.*
- 79: em_estro5.*
- 80: em_estro6.*
- 81: em_estro7.*
- 82: em_estro8.*
- 83: em_estro9.*
- 84: em_estro10.*
- 85: em_estro11.*
- 86: em_estro12.*
- 87: em_estro13.*
- 88: gb_gss1.*
- 89: gb_gss2.*
- 90: gb_gss3.*
- 91: gb_gss4.*
- 92: em_gss1.*
- 93: em_gss2.*
- 94: em_gss3.*
- 95: em_gss4.*
- 96: gb_gss5.*
- 97: gb_gss6.*
- 98: gb_gss7.*
- 99: gb_gss8.*
- 100: gb_gss9.*
- 101: em_gss5.*
- 102: em_gss6.*
- 103: em_gss7.*
- 104: em_gss8.*
- 105: em_gss9.*
- 106: em_gss10.*
- 107: em_gss11.*
- 108: gb_gss10.*
- 109: gb_gss11.*
- 110: em_gss12.*
- 111: gb_gss12.*
- 112: gb_gss13.*
- 113: gb_gss14.*
- 114: gb_gss15.*
- 115: gb_gss16.*
- 116: gb_gss17.*

Result No.	Score	Query Match	Length	ID	Description
C 1	564.8	39.7	751	AI660567	w68c03.x
C 2	554	39.0	695	B407909	BE407909
C 3	515.8	36.3	522	BE464528	BE464528
C 4	411.4	28.9	530	BE198430	BE198430
C 5	411.4	28.9	530	BE198430	BE198430
C 6	370.8	26.0	440	AW463722	AW463722
C 7	356.2	25.0	435	AA112513	AA112513
C 8	336.2	23.6	428	AA112513	AA112513
C 9	321.6	22.6	462	AA113289	AA113289
C 10	301.4	21.2	398	AA1368245	AA1368245
C 11	234.6	16.5	491	F22931	F22931
C 12	193.2	13.6	529	AW857217	AW857217
C 13	187.8	13.2	545	B2283475	B2283475
C 14	184	12.9	420	AW935681	AW935681
C 15	164	11.5	258	AW758472	AW758472
C 16	162.4	11.4	193	AW796557	AW796557
C 17	158	11.1	635	AI388785	AI388785
C 18	142	10.0	248	AI801220	AI801220
C 19	130.4	9.2	469	BE061124	BE061124
C 20	118.2	8.3	379	C69305	C69305
C 21	113.8	8.0	639	AL110451	AL110451
C 22	112	7.9	820	AV398238	AV398238
C 23	105.4	7.4	548	BE201648	BE201648
C 24	105	7.4	360	D88728	D88728
C 25	80.6	5.7	911	AL262733	AL262733
C 26	80.4	5.7	289	C12082	C12082
C 27	77	5.4	360	FR0018487	FR0018487
C 28	74.2	5.2	179	AL011383	AL011383
C 29	66.4	4.7	925	AL033013	AL033013
C 30	65	4.6	574	AL011361	AL011361
C 31	64.6	4.5	932	AL056742	AL056742
C 32	59.6	4.2	935	AL056051	AL056051
C 33	58.6	4.1	935	AL056051	AL056051
C 34	57.2	4.0	925	CNS0091P	CNS0091P
C 35	53.6	3.8	197	AW268319	AW268319
C 36	53.6	3.8	201	AI345656	AI345656
C 37	53.6	3.8	218	AI344839	AI344839
C 38	53.6	3.8	218	AI345593	AI345593
C 39	53.6	3.8	218	AW302967	AW302967
C 40	53.6	3.8	219	AW301388	AW301388
C 41	53.6	3.8	239	AW268166	AW268166
C 42	53.6	3.8	251	AI344947	AI344947
C 43	53.2	3.7	1101	CNS017SY	CNS017SY
C 44	52	3.7	202	AW301448	AW301448
C 45	52	3.7	207	AW301461	AW301461

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

LOCUS AI660567 751 bp mRNA EST 18-DEC-1999
DEFINITION w68c03.x1 Soares.Dieckgraefe.colon_NHCD Homo sapiens CDNA clone IMAGE:2346244 3' similar to TR:Q13193 Q13193 CBL-B TRUNCATED FORM 2. ; mRNA sequence.
ACCESSION AI660567.1 GI:4764137
VERSION AI660567.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 751)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 864 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 458.
Location/Qualifiers
1. .751
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2346244"
/clone_lib="Soares.Dieckgraefe.colon_NHCD"
/tissue_type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: p7T3B-Pac (Pharmacia) with a modified polylinker; Site: 1; Not I; Site: 2; Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3', TGTTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieckgrm.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 124 a 222 c 228 g 172 t 5 others
ORIGIN

Query Match 39.7%; Score 564.8; DB 12; Length 751;
Best Local Similarity 94.8%; Pred. No. 1.5e-120;
Matches 614; Conservative 0; Mismatches 32; Indels 3; Gaps 3;
Qy 775 ggcagttacatcttcggcccgagctgactcgctggggcagtgaggccatcggtatgtg 834
Db 751 GGCAGTTACATCTTCGGCCCAACTGACTCGTTGGGGCAGTGGCCATCGGTATGTG 692
Qy 835 agctcagatggcagcatcttcggagaccatccctgcacacacacccctgtctcc-aggtgct 893
Db 691 AGCTCAGATGGCAGCATCTTGCAGACATCTTGGCCAAACACNCCTGTTCCCAAGTGCT 632
Qy 894 cctggaggacacagagcagcgtcttaccctaccagatggaaagaccacacaccaga 953
Db 631 CTGAAGGGACAGAGGAGCGGTCTTCA-CTCTACCCAGATGGAAAGACACACACCCAGNN. 573
Qy 954 cctgactgagctggccaggcagacaccccgagcgcacatccactgtgtcagagagagct 1013
Db 572 CCTGACTGAGTCTCGCCAGGACAGAA-CCCAGCAGCGCATCCAGGTGTGAGAGGAGCT 514
Qy 1014 gcactctactgggcagctgactccacatttgagctctgcaagatctgtgctgagagcaa 1073

ALIGNMENTS

Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 508.

FEATURES

source
 1. 522
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3180165"
 /clone_lib="NCI_CGAP_Lu24"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 117 a 168 c 153 g 84 t
 ORIGIN

Query Match 36.3%; Score 515.8; DB 35; Length 522;
 Best Local Similarity 99.6%; Pred. No. 3.1e-109;
 Matches 517; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 866 ctgccacaacccctgtccagggtctctggagggaagaaggcgctttactctt 925
 DB 4 CCGCCAAACACCCCTGCTCCAGGTGCTCTGGAGGACAGAGGACGCTTACTCTT 63
 QY 926 accagatggaagaccacacacccagacctgactgagctcgccaggcagacccaccg 985
 DB 64 ACCAGATGGAAGACCCACACCCAGACCTGACTGAGCTCGCCAGGAGAACCCACG 123
 QY 986 agcgatccactgtcagaggagcagctgcagctctacttggccatgggactccacattg 1045
 DB 124 AGCGCATCCACGTGTCAGAGGAGCAGTGCAGCTCTACTGGGCATGGACTCCACATTG 183
 QY 1046 agctctcaagatctgtctgagagaacaagatgtgaagattgagcctgcggccacc 1105
 DB 184 AGCTCTCAGAGATCTGTCTGAGAGCAACAGGATGTGAGATTGAGCTGCGGGCAC 243
 QY 1106 tgcctcagactgctgctgctgctggcagcaactcgacagccagacctgccctctt 1165
 DB 244 TGCTCTCAGCTGCTGCTGGCTGGCTGGCAGCACTCGGACAGCCAGACTGCCCTTCT 303
 QY 1166 gcgcctcagagatcaagggtcgaggccgctgagtatctaccagttccaggtcaggcta 1225
 DB 304 GCCGCTCGAGATCAAGGGGTGGAGGCGGTGAGTATCTACCACTTCCACGGTCAAGCTA 363
 QY 1226 ctgctgaggactcagggaacagcagtgaccagggaagcaggagtgaggctggggcagg 1285
 DB 364 CTGCTGAGGACTCAGGGAACAGCAGTGACACAGGAGCAGGAGTTGGAGCTGGGGCAG 423
 QY 1286 tgcctcttggctctcatttgcctccacagggccagatctgccccccaggaagcccgaa 1345
 DB 424 TGCCCCCTTTGGCTCTCTCCATTCGCCCCACGGCCAGATTCGCCCCCATGAAGCCAGAA 483
 QY 1345 atgcccagccgaagtggagactcctaaagggaactccc 1384
 DB 484 ATGCCAGCGGAAGTGAGACTCTTAAGGGGAGACTCCC 522

RESULT 4

BE198430 LOCUS 533 bp mRNA EST 26-JUN-2000
 DEFINITION ug78e05.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone IMAGE:1548512 5' similar to gb:X57110 PROTO-ONCOGENE C-CBL (HUMAN)>. mRNA sequence.
 ACCESSION BE198430
 VERSION BE198430.1 GI:8710599
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 REFERENCE 1 (bases 1 to 533)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:953860
 Seq primer: -40RP from Gibco
 High quality sequence stop: 466.
 FEATURES
 Location/Qualifiers
 1. 533
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:1548512"
 /sex="female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

ORIGIN 125 a 169 c 133 g 106 t

Query Match 28.9%; Score 411.4; DB 34; Length 533;
 Best Local Similarity 85.7%; Pred. No. 4.3e-85;
 Matches 457; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 634 gacgtttccacaggctcttttccagccatggccacacactctcctcgaactgcgactctg 693
 DB 1 GACGTCTTCACAGGCTCTTTACGCCGTGGCCACTTTACTGAGGATTTGGCAACTCTG 60
 QY 694 gcagtcaaccccccagctacatgctctcctcactatgatgagtcctcgaagcgctctg 753
 DB 61 GCTGTACACCATCTGCTGTACATGGCCTTCTTACCTACCATGAGTTCGAAACACGCTG 120
 QY 754 caggcctgcaggggacacagccaggcagttacatcttcggcccgctgtactgcctcgggg 813
 DB 121 CAGGCTACAGGGACAAACAGGAGGAGCTATCTTCGGCCAAAGCTATCCCGCTGGGG 180
 QY 814 cagtgggccaatcggtctatgagctcagatggcagctcctcagaccatccctgcgaac 873
 DB 181 CAGTGGGCAATTTGATACGTGAGTCCGATGGAGAGTCTCTGAAACCATCTCTCTCAAC 240
 QY 874 aaacccctgtcccagggtgctcctggaggacagaagcggctctcactctaccagat 933
 DB 241 AAACCTCTGCTGCAAGTGTCTCTGAAGGGACAAAAGGAGCGCATCTCTCTCTCCCTGAT 300
 QY 934 ggaagaccctcaacccagacctgactgagctggccaggcagaaccccgagcgcatc 993

Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806.

```

BASE COUNT      100 a 142 c 120 g      78 t
ORIGIN

Query Match      25.1%; Score 370.8; DB 22; Length 440;
Best Local Similarity 90.4%; Pred. No. 1e-75;
Matches 396; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 748 cgtctgcaggctgcaggacacccagcaggcagttacatctccggcccgactgtactcgc 807
      |||||
Db 1 CGTCTGCAGACTTCAGAGACACAGCCAGGCGAGCTACATCTCCGGCCAGCTGCACCTGC 60

Qy 808 ctggggcagtgggccatctggctatgtgagctcagatggcgagcctctctgagacacccct 867
      |||||
Db 61 CTGGGGAAGTGGGCGCATCGGATAGCTGAGCTCAGACGGCAGCATCTTGCAGACCATCCCT 120

Qy 868 gccacaacacccctgtcccagggtctcctggaggagcagagcagcgtcttaccctctac 927
      |||||
Db 121 CTCACAAACCCCTGTTCCAGGTCCTCTCGAAGGACAGAGAGGCTCTACCTCTAC 180

Qy 928 ccagatgaaagacccacacccagacactgactgagctcgtggcaggcagacaccccgagc 987
      |||||
Db 181 CCAGATGGGAAGAACCAACACACCCGAGCTGACCGAATCTGCCACATGGAAACGCGATCAG 240

Qy 988 cgcattcagctgtcagaggagcagctcagctctactggccgagcagcagcagcattgag 1047
      |||||
Db 241 CACATCCACGTGTCGGAGGAGCAGCTGACGCTCTACTGGGCCATGATGACTCTAGCTTTGAG 300

Qy 1048 ctctcgaagtctgtctgtagagcaacaaggatgtgaagattgagccgtgctggcgacctg 1107
      |||||
Db 301 CTCTGCAAGATCTGTGCCGAGAGCAACAGGACGTGAAGATAGAGCCCTGCGGGCACCTG 360

Qy 1108 ctctgagactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1167
      |||||
Db 361 CTCTGTAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420

Qy 1168 cgtctgcagatcaagggc 1185
      |||||
Db 421 CGCTGCAGATCAAGGCG 438

RESULT 7
AAL12513 435 bp mRNA EST 23-DEC-1997
LOCUS zm28c07.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
DEFINITION IMAGE:526956 5' similar to gb:X57110 PROTO-ONCOGENE C-CBL (HUMAN);,
mRNA sequence.
ACCESSION AAL12513
VERSION AAL12513.1 GI:1665062
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 435)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaaskis,E.,
Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
Contact: Wilson RK
97044478
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

```

WARNING: There is evidence that suggests that the 384-well parent plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 336.

FEATURES

Location/Qualifiers
1..435
/organism="Homo sapiens"
/db_xref="GDB:3918385"
/db_xref="taxon:9606"
/clone="IMAGE:526956"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site: 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Pancreatic adenocarcinoma cell line. Average insert size: 1.0 kb. Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."

BASE COUNT 94 a 136 c 121 g 82 t 2 others
ORIGIN

Query Match 25.0%; Score 356.2; DB 1; Length 435;
Best Local Similarity 94.1%; Pred. No. 2.5e-72;
Matches 381; Conservative 0; Mismatches 20; Indels 4; Gaps 1;

```

Qy 763 agggacaagccaggcagttacatcttcggcccgagctgactcgtcgtggcgagtgggcc 822
      |||||
Db 1 AGGGACAAGCCAGGCGAGTTACATCTTCGGCCCGAGCTGACTCGCTGGGNCAGTGGGCC 60

Qy 823 atcggtatgtgaactcagatggcagatcctcgcagacatcctcgcagacatcctcgcacacacccctg 882
      |||||
Db 61 ATCGNTATTGTGAGCTCAGATGGGAGCATCTGCGAGACCATCCCTGCCAACAAACCCCTG 120

Qy 883 tccagggtcctcgtggaggacagaagcagcgttctacctctaccagatggaaagacc 942
      |||||
Db 121 TCCAGAGTGTCTCTGGAGGGACAGAGAGCGGCTCTTACCTTACCCAGATGGAAGAAC 180

Qy 943 cacaacccagacctgactgactgagctcgtggccagcagaaacccagcagcattccagtgatca 1002
      |||||
Db 181 CACAACCCAGACCTTGACTGAGCTCGGCCAGGACAGAACCCAGCAGCATCCAGTGTCAT 240

Qy 1003 gaggagcagctgcagctctacttggcccgatggactccacatttgagctctcgaagatctgt 1062
      |||||
Db 241 GAGGAGCAGCTGCAGCTCTACTGGCCATGGACTCCACATTTGAGCTCTCTCAAGATCTGT 300

Qy 1063 gctgagagcaacaaggatgtgaagattgagcctcgtggcgagcagcagcagcagcagcagcagc 1122
      |||||
Db 301 GCTGAGAGCAACAAGGATGTGAAGATTGAGCCGTGCGGCGACCATCTGCTCGACGCTGCTC 360

Qy 1123 ctggctgctggcgagca----ctcgagacagcagcagcagcagcagcagcagcagcagcagcagc 1163
      |||||
Db 361 CTGGCTGCTGGCAGAACTTCGGACAGCCAGACCTTGCCCT 405

```

RESULT 8

```

AAL12513 428 bp mRNA EST 07-JUL-2000
LOCUS AW426122
DEFINITION 59653 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW426122
VERSION AW426122.1 GI:6954069
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 428)

```

AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.F., Grosse,W.M.,
Bennett,G.A., Fahrnenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
W.W. and Keesle,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT USDA, ARS, US Meat Animal Research Center
Contact: Smith TPL
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithem@mail.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACAT
BACKWARD: GTTTCCTCCATGACGAG
Plate: 27 row: 1 column: 8
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1. 428
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 480V"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos"
BASE COUNT 102 a 133 c 121 g 72 t
ORIGIN

Query Match 23.6%; Score 336.2; DB 22; Length 428;
Best Local Similarity 89.4%; Pred. No. 1e-67; Indels 0; Gaps 0;
Matches 362; Conservative 0; Mismatches 43;
QY 697 gtaacacacacaggtacatgccccttcctacattatgatgaggtccaaagaggtctgcag 756
Db 24 GGCACACACCGGGTACATGGCTTCCTCACATAGATGAGGTCCAAACCGCTGTGCAG 83
QY 757 gctgcagagacacgacgagcaggtacattcttcgcccagctgtactgcctggggcag 816
Db 84 ACTTTCAGAGACAGCAGCAGCAGTACATCTTCGGCCACAGTGCATCGCTGGGGAG 143
QY 817 tggggccatggctatgtgagctagatggcagcatcctgcagaccatccctgccacaaa 876
Db 144 TGGGCCATCGATACGTGAGCTCAGACGGCAGCATCTTCAGACCATCCCTCTCAACAA 203
QY 877 cccctgtcccagctgtctggaggacagagagcgtttctaccttaccacagatgga 936
Db 204 CCCCTGTTCAGGTGCTCTGGAGGACAGAGAGGAGGTTTCTACTCTACCCAGATGG 263
QY 937 aagaccac 996
Db 264 AAGAACACACACCGGACCTGACCGAATCTGCCACATGGAACCGCATCAGCATCCAC 323
QY 997 gtctcagagagagctgagctctactggtccatgagctcattgacattgagctgcag 1056
Db 324 GTGTGGAGAGAGAGTGTGAGCTTACTGGGCGATGGACTTACGTTTGGAGTCTGCAAG 383
QY 1057 atctgtgtgagagacacagaggtgtgagattgagcggcgagg 1101
Db 384 ATCTGTGCCAGAGACACAGAGGAGCTGAAGATAGACCCCTCGGG 428

FEATURES source
1. 428
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 480V"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos"
BASE COUNT 102 a 133 c 121 g 72 t
ORIGIN

ACCESSION AA113289.
VERSION AA113289.1 GI:1664994
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 462)
REFERENCE Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasis,E.,
Underwood,K., Wohldmann,P., Waterston,K., Wilson,R. and Warr,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
JOURNAL MEDLINE
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
WARNING: There is evidence that suggests that the 384-well parent
plate of this clone contains both human and mouse derived clones.
Thus, the origin of this clone is uncertain. This caution should be
kept in mind should you use this clone.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 367.
Location/Qualifiers
1. 462
/organism="Homo sapiens"
/db_xref="GDB:3918385"
/db_xref="taxon:9606"
/clone="IMAGE:526956"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT 76 a 141 c 137 g 99 t
ORIGIN
Query Match 22.6%; Score 321.6; DB 1; Length 462;
Best Local Similarity 94.6%; Pred. No. 2.5e-64;
Matches 349; Conservative 0; Mismatches 18; Indels 2; Gaps 2;
QY 1054 aagatctgtgtgagacacacagatgtgaagattgagcgtgcgggcaacctgtctgc 1113
Db 462 AAGATCTGTGTGAGACACACAGATGTGAAGATTGAGCGTCGGGCGACCTTTCTGC 403
QY 1114 agctgtctccctggctcctggcagacctcgacag-cacagacctccctcttcccgctg 1172
Db 402 AGCTGTCTGCTGGNTGCTGGCAGCATTCGACAGCCAGACCTTCCCTTCTGCCGCTG 343
QY 1173 cgagatcaaggtcgtggggagcgcgtgagtattctaccagttccacggtcaggctactgtga 1232
Db 342 CGAGATCAAGSGCTGGGAGCGCGTGAGTATCTACAGTTCTACGGTCAAGGTTACTGTGA 283
QY 1233 ggactcaggggaacagcagctgaccacgggaagcaggaggttgagctggggcagggtgcacct 1292
Db 282 GGACTTCAGGGAACACAGCAGTACAGGAGGAGGAGTGGAGCTGGGCGAGGTGCCCT 223
QY 1293 ttccgctcctccattgccccccacagccagatctgccccccaggaagccacgaatgccca 1352

FEATURES source
1. 462
/organism="Homo sapiens"
/db_xref="GDB:3918385"
/db_xref="taxon:9606"
/clone="IMAGE:526956"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTGAGTTTCTTTTCTTTTCTTTT 3'"

RESULT 9
AA113289/c 462 bp mRNA EST 23-DEC-1997
LOCUS zM28c07.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
DEFINITION IMAGE:526956 3' similar to TR:G862411 G862411 CEL-B TRUNCATED FORM
2.; mRNA sequence.

```

Db 222 TTNGGCTCTCCATTTGCCCCACGGCCAGATCTGCCCCCGAGGAAGCCAGAAATGCCCA 163
QY 1353 gccaaagtgaactcctaaggggaactccctccagctgcctggagcccgagacc 1412
Db 152 GCCAAAGTGAAGTCTTAAAGGGGAAT-CCNCNCTNTGGCTGGAGCCCGCAGGACN 104
QY 1413 tgcccccgcc 1421
Db 103 NCTNCCGC 95

RESULT 10
AI368245/c
LOCUS
DEFINITION QW16409.x1 NCI-CGAP_Ut3 Homo sapiens cDNA clone IMAGE:1991273 3'
similar to TR:Q13193 Q13193 CBL-B TRUNCATED FORM 2.; contains
PTR5.t3 MER22 repetitive element.; mRNA sequence.
ACCESSION AI368245
VERSION AI368245.1 GI:4137990
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 398)
AUTHORS Contact: Robert Strausberg, Ph.D.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Sequencing Arrayed by: Greg Lennon, Ph.D.
CDNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/obrp/image/image.html
Insert Length: 648 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 359.
FEATURES
source
1..398
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1991273"
/clone_lib="NCI-CGAP_Ut3"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="PH103"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"
BASE COUNT 61 a 125 c 129 g 83 t
ORIGIN

Query Match 21.2%; Score 301.4; DB 10; Length 398;
Best Local Similarity 99.7%; Pred. No. 1.1e-59;
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1120 tgcttgctgctggcagcactcgagccagccagcactgccccttgcgcgtgcagacc 1179
Db 398 TGCTTGCTTGCTGGCAGCAGCTGGACAGCAGCAGCTGCGCTTGCGCGCTGCCAGATC 339
QY 1180 aagggtggagggccttgagtattaccagttccagcgttcaggctactgctgagactca 1239
Db 338 AAGGGTGGAGGGCCGTGAGTACTACCAAGTTCGCCGGTCAGGCTACTGCTGAGGACTCA 279
QY 1240 gggaacagcagtgaccaggaaggcagggtgaggctggcaggtggccctttcggct 1299

```

```

Db 278 GGGACACAGTGTACCAGAGGAGGAGTGGAGCTGGGCGAGGTGCCCTTCGGCT 219
QY 1300 cctccatgccccccagcggcagatctgccccccaggaagccccagaatgccagccgaaa 1359
Db 218 CTTCCATTGCCCCACGGCCAGATCTGCCCCCGAGGAAGCCAGAAATGCCAGCCGAAA 159
QY 1360 gtgagactcctaaagggaactccctccagctgcctgggagccccagaccctgcccg 1419
Db 158 GTGACACTCCCTAAGGGGAATCTCCCTCCAGTGGCTGGAGCCCGGAGGACCTTGCCCG 99
QY 1420 gcc 1422
Db 98 GGC 96

RESULT 11
F22931
LOCUS
DEFINITION SSC17A01 Porcine small intestine cDNA library Sus scrofa cDNA clone
cl7a01 5', mRNA sequence.
ACCESSION F22931
VERSION F22931.1 GI:2578544
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 491)
AUTHORS Wintero,A.K., Fredholm,M. and Davies,W.
TITLE Evaluation and characterization of a porcine small intestine cDNA
library: analysis of 839 clones
JOURNAL Mamm. Genome 7 (7), 509-517 (1996)
MEDLINE 96327607
COMMENT Contact: A.K. Wintero
Department of Animal Science and Animal Health, Division of Animal
Genetics, The Royal Veterinary and Agricultural University
Bulowsvej 13, 1870 Frederiksberg C, Denmark.
FEATURES
source
1..491
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="cl7a01"
/clone_lib="Porcine small intestine cDNA library"
/note="directionally cloned cDNA in XLI-blue MRF"
BASE COUNT 116 a 142 c 145 g 79 t 9 others
ORIGIN

Query Match 16.5%; Score 234.6; DB 37; Length 491;
Best Local Similarity 82.0%; Pred. No. 3.3e-44;
Matches 282; Conservative 0; Mismatches 60; Indels 2; Gaps 1;

QY 832 gtgagctcagatggcagcactctgcagaccatccctgcccaacaaacctgtcccggtg 891
Db 1 GTGAGCTCAGATGGCAGCATCTGCAGACCATCCCTCTCAACAACCTCTGTTCAGCA 60
QY 892 ctcctggaggacagaaggagcgttcttactctaccatggaagaaagaccccaacca 951
Db 61 CTCCTGGAAGGACAAAAGGAGGCTTCTACTCTACTACCTGATGGGAAGACCAACACCG 120
QY 952 gacctgactagctcgcccgaggcagaaccccgagcgctccacgtgtcagaggagcag 1011
Db 121 GACCTGACGGAGCTCTCGTAACGGGACCTATCAACGCATCCAGCTGTGGAGGAGCAG 180
QY 1012 ctgcagctctactgggcatggactccacatttgagctctgcaagatctgtctgagagc 1071
Db 181 CTGAGCTGTACTGGGCGCATGAATCCACATACGAGCTGTGCAAGATCTGTGCCGAGAGA 240
QY 1072 aacagcagtggaagattgagcgtgcgggcaacctgtctgtcgag--ctgtgctggtgctg 1129
Db 241 AACAGGACGTGAAGATTGAGCCATGCGGGGCACTGCTCTGCAAGCCCGCTGCTCGGCTA 300

```

[illegible]

/db_xref="taxon:9606"
/clone="IMAGE:2278445"
/clone_lib="NCI_CGAP_Ut3"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"

BASE COUNT 37 a 80 c 85 g 56 t
ORIGIN

Query Match 11.5%; Score 164; DB 12; Length 258;
Best Local Similarity 100.0%; Pred. No. 6.4e-28;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1259 aaggcaggagttgagctggggcagggtgcccccttcggtctctccattgccccacggc 1318
DB 258 AAGGCAGGAGTTGAGCTGGGGCAGGTGCCCTTTTCGCTCTCTCCATTGCCCCACGGC 199
QY 1319 cagatctgccccccagggaagccagaaatgcccagccgaaagtgagactcctaaagggga 1378
DB 198 CAGATCTGCCCCCCCAGGAAGCCAGAAATGCCCGCCGAAAGTGAGACTCTCTAAAGGGGA 139
QY 1379 actccccctcagctggtggtgggacccccagaccctgccccggcc 1422
DB 138 ACTCCCTCCAGCTGGCTGGGACCCAGGACCCTGCCCGGCC 95

Search completed: December 7, 2000, 12:41:46
Job time: 24589 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2000, 05:51:57 ; Search time 1490.04 Seconds
(without alignments)
6419.158 Million cell updates/sec

Title: US-09-434-708-1
Perfect score: 1547
Sequence: 1 cgcgaggtccctgctct.....caaaaaaaaaaaaaaaaaa 1547

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est25:*
- 26: gb_est26:*
- 27: gb_est27:*
- 28: gb_est28:*
- 29: gb_est29:*
- 30: gb_est30:*
- 31: gb_est31:*
- 32: gb_est32:*
- 33: gb_est33:*
- 34: gb_est34:*
- 35: gb_est35:*
- 36: gb_est36:*
- 37: gb_est37:*
- 38: gb_est38:*
- 39: gb_est39:*
- 40: gb_est40:*
- 41: em_estba:*
- 42: em_estfun:*
- 43: em_esthum1:*

- 44: em_esthum2:*
- 45: em_esthum3:*
- 46: em_esthum4:*
- 47: em_esthum5:*
- 48: em_esthum6:*
- 49: em_esthum7:*
- 50: em_esthum8:*
- 51: em_esthum9:*
- 52: em_esthum10:*
- 53: em_esthum11:*
- 54: em_esthum12:*
- 55: em_esthum13:*
- 56: em_esthum14:*
- 57: em_esthum15:*
- 58: em_esthum16:*
- 59: em_esthum17:*
- 60: em_esthum18:*
- 61: em_esthum19:*
- 62: em_esthum20:*
- 63: em_estin1:*
- 64: em_estin2:*
- 65: em_estin3:*
- 66: em_estin4:*
- 67: em_estom:*
- 68: em_estov1:*
- 69: em_estov2:*
- 70: em_estpl1:*
- 71: em_estpl2:*
- 72: em_estpl3:*
- 73: em_estpl4:*
- 74: em_estpl5:*
- 75: em_estro1:*
- 76: em_estro2:*
- 77: em_estro3:*
- 78: em_estro4:*
- 79: em_estro5:*
- 80: em_estro6:*
- 81: em_estro7:*
- 82: em_estro8:*
- 83: em_estro9:*
- 84: em_estro10:*
- 85: em_estro11:*
- 86: em_estro12:*
- 87: em_estro13:*
- 88: gb_gss1:*
- 89: gb_gss2:*
- 90: gb_gss3:*
- 91: gb_gss4:*
- 92: em_gss1:*
- 93: em_gss2:*
- 94: em_gss3:*
- 95: em_gss4:*
- 96: gb_gss5:*
- 97: gb_gss6:*
- 98: gb_gss7:*
- 99: gb_gss8:*
- 100: gb_gss9:*
- 101: em_gss5:*
- 102: em_gss6:*
- 103: em_gss7:*
- 104: em_gss8:*
- 105: em_gss9:*
- 106: em_gss10:*
- 107: em_gss11:*
- 108: gb_gss10:*
- 109: gb_gss11:*
- 110: em_gss12:*
- 111: gb_gss12:*
- 112: gb_gss13:*
- 113: gb_gss14:*
- 114: gb_gss15:*
- 115: gb_gss16:*
- 116: gb_gss17:*

117: gb_gss18:*
118: gb_gss19:*
119: em_gss13:*
120: gb_gss20:*
121: gb_gss21:*
122: gb_gss22:*
123: gb_gss23:*
124: gb_gss24:*
125: em_gss14:*
126: em_gss15:*
127: em_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	668.8	43.2	751	AT1660567	we8c03.x
2	554	35.8	695	BE407909	601299249
3	515.8	33.3	522	BE464528	hv84hll.x
4	512	33.1	530	AT1690580	tq02ail.x
5	411.4	26.6	533	BE198430	ug78e05.y
6	396.4	25.6	398	AT1368245	qm16f09.x
7	393.6	25.4	462	AA113289	zm28c07.s
8	370.8	24.0	440	AA113289	zm28c07.s
9	356.2	23.0	435	AA112513	zm28c07.r
10	356.2	21.7	428	AA112513	zm28c07.r
11	258	16.7	258	AT1758472	59553.MAR
12	236	15.3	248	AT1801220	to86c10.x
13	234.6	15.2	491	F22931	SSC17A01.PO
14	193.2	12.5	529	AW857217	RC3-C7030
15	187.8	12.1	545	BE283475	601102508
16	184	11.9	420	AW935581	RC3-D7001
17	166.6	10.8	251	AT144947	tblc001.x
18	165	10.7	197	AT144947	tblc001.x
19	163	10.7	218	AT144839	ta99g09.x
20	163	10.7	219	AT144839	ta99g09.x
21	163.4	10.6	207	AW301461	xs77c05.x
22	162.4	10.5	193	AW301388	xs76c04.x
23	158	10.2	635	AW796557	CM4-UM003
24	157	10.1	218	AT302967	xr87c06.x
25	155.4	10.0	218	AT345593	tb78d02.x
26	152.2	9.8	201	AT345556	tb78f04.x
27	150.6	9.7	202	AW301448	xs77a09.x
28	149	9.6	239	AW268166	xr93a08.x
29	130.4	8.4	469	BE061124	OY0-B7C04
30	118.2	7.6	379	C69305	C69305
31	113.8	7.4	639	AL110451	DKE2P434P
32	112	7.2	820	AV398238	AV398238
33	105.4	6.8	548	BE201648	fk96d07.x
34	105	6.8	360	D68728	CELK055AVF
35	80.6	5.2	911	CNS03VPW	AL262733
36	80.4	5.2	289	40	CELK055AVF
37	77	5.0	360	C12082	CELK055AVF
38	74.2	4.8	179	FR0018487	CELK055AVF
39	66.4	4.3	925	CNS0091P	CELK055AVF
40	65	4.2	574	FR0018465	CELK055AVF
41	64.6	4.2	932	CNS0072Q	CELK055AVF
42	59.6	3.9	121	CNS006XK	CELK055AVF
43	59.2	3.8	925	CNS0091P	CELK055AVF
44	59	3.8	925	CNS006XK	CELK055AVF
45	54.2	3.5	932	CNS0072Q	CELK055AVF

ALIGNMENTS

LOCUS AT1660567 751 bp mRNA EST 18-DEC-1999
DEFINITION we8c03.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone IMAGE:2346244 3' similar to TR:Q13193 Q13193 CBL-B TRUNCATED FORM 2. /, mRNA sequence.
ACCESSION AT1660567
VERSION AT1660567 1 GI:4764137
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 751)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 864 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 458.
FEATURES
source
1. 751
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2346244"
/tissue_type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pTR3D-pac (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' GTTACCAATCTGAATGGAGCGCGCGCGCTTTTCTTTTCTTTT 3'], TGTACCAATCTGAATGGAGCGCGCGCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR3 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieckim.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 124 a 222 c 228 g 172 t 5 others
ORIGIN

Query Match 43.2%; Score 668.8; DB 12; Length 751;
Best Local Similarity 95.4%; Pred. No. 3.5e-146;
Matches 718; Conservative 0; Mismatches 32; Indels 3; Gaps 3;
Qy 787 ggcagttacattccggccagctgtactcgcctggcgagtgccatcgactgtg 846
|||||
Db 751 GGCAGTTACATCTTCGCGCCCAACTGACTGCTTTGGGGCAGTGGCCATCGGCTATGTG 692
Qy 847 agctcagatggcagcatctctgcagaccatccctgcgcacacacccctgtccc-agggtgt 905
|||||
Db 691 AGCTCAGATGGCAGCATCTTGCAGACATCCTTGGCCCAACACNCCTTGTCCCAAGGTGCT 632
Qy 906 cctggaggagcagagcagcggcttctaccctaccagatggaagaccacacaccaga 965
|||||
Db 631 CTGAAGGGGACAGAGGGAGCGGCTTCTA-CTCTACCCAGATGGAAAGACCAACACCCAGNN-573
Qy 966 cctgactgagctcgccaggcagcagaaccccgagcgcgcattccacgtgtcagagagcagct 1025
|||||
Db 572 CCGTACGTGAGCTCGCCAGCAGAGAA-CCCAGCAGCGCATCCACGTGTCAGAGGAGCAGCT 514
Qy 1026 gcagctctactggccagtgactccacatttgagctctgcaagatctgtgctgagagcaca 1085
|||||

513 GCAGCTCTACTGGCCATGACCTCCACATNTGAGCTCTGCAAGATCTGTGCTGAGAGCAA 454
QY 1086 caaggatgtgaagattgagcctggtgggacacgtgtctgacgtctgctgctgctgctg 1145
Db 453 CAAGGATGTGAAGATTGAGCGGTGGGACACCTGTCTGACGTGCTGCTGCTGCTGCTG 394
QY 1146 ggaactcggacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1205
Db 393 CGAGCACTGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 334
QY 1206 cgtgagtattcattcagctgagcagcagcagcagcagcagcagcagcagcagcagc 1265
Db 333 CGTGAATCTACCACTGACAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 274
QY 1266 caaggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1325
Db 273 CCAGGAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 214
QY 1326 acggccagatctcctcccccaggaagccaggaagccaggaagccaggaagccaggaag 1385
Db 213 ACGCCAGATCTGCCCCCGAGGAAGCCAGAAATGCCAGGCGAAGAGTGAGACTCTAAA 154
QY 1386 ggggaactccctccagctgctgctgctgctgctgctgctgctgctgctgctgctgct 1445
Db 153 GGGGAATCTCCCTCCAGCTGCGCTGGGAGCCAGGAGCCAGGAGCCAGGAGCCAGGAG 94
QY 1446 gcaccagatgtgctgctcaaggagagcccaaggagcagcagcagcagcagcagcagc 1505
Db 93 GCACCAGATGCTGCTCAAGGAGGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 34
QY 1506 taaactgcaagcctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1538
Db 33 TAAACTGCCAAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1

RESULT 2
BE407909 695 bp mRNA EST 21-JUL-2000
LOCUS 501299249F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3629101 5',
DEFINITION mRNA sequence.
ACCESSION BE407909
VERSION BE407909.1 GI:9344359
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 695)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: LCM315 row: b column: 14
High quality sequence stop: 686.
FEATURES
1..695
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3629101"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; vector: pORF7; Site1: XhoI;
Site2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA Synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 118 a 241 c 206 g 130 t
ORIGIN

Query Match 35.8%; Score 554; DB 35; Length 695;
Best Local Similarity 94.0%; Pred. No. 2.4e-119;
Matches 643; Conservative 0; Mismatches 30; Indels 11; Gaps 6;
QY 75 agtcagatgtgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 134
Db 1 AGTCAGGATGTGTCAGCGCCTAGAGACCAATGCTGACACCCCGGCTGTCCTGAGTCC 60
QY 135 ccttcgtgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 193
Db 61 CCTTCGCTGGGACCTGCTGCCCGCACAGCGACAGCTGCTTCGACAGAGTGCCCAT 120
QY 194 ctccgctggcctggcctggcctggcctggcctggcctggcctggcctggcctggcctggc 251
Db 121 CTGCGCGGCG 180
QY 252 actcat--ctacctggcctgagcagcagcagcagcagcagcagcagcagcagcagcagc 309
Db 181 ACTCATTTCTCTCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTC 240
QY 310 cccgggctggcctggcctggcctggcctggcctggcctggcctggcctggcctggcctgg 365
Db 241 CCGCGGCG 300
QY 366 gcgacagctggcctggcctggcctggcctggcctggcctggcctggcctggcctggcctgg 424
Db 301 CGGACTGCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTC 360
QY 425 tcttcctggcctggcctggcctggcctggcctggcctggcctggcctggcctggcctggc 484
Db 361 TCTTCCCG 420
QY 485 ccttcctggcctggcctggcctggcctggcctggcctggcctggcctggcctggcctgg 544
Db 421 CCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480
QY 545 tcttcctggcctggcctggcctggcctggcctggcctggcctggcctggcctggcctggc 604
Db 481 TCTTGGGACCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTG 540
QY 605 ttgacctcagcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 664
Db 541 TTGACCTCAGCTGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
QY 665 ttgacctcagcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 724
Db 601 TTCACCCATGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTGTCCTG 659
QY 725 acatggccttcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 748
Db 660 ACATGGCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 683

RESULT 3
BE464528 522 bp mRNA EST 27-JUL-2000
LOCUS hv84hl1.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3180165 3'
DEFINITION similar to TR:Q13193 Q13193 CBL-B TRUNCATED FORM 2.; mRNA
sequence.
ACCESSION BE464528
VERSION BE464528.1 GI:9510303
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 522)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
 Seq primer: -40UP from Gibco
 High quality sequence stop: 508.
 Location/Qualifiers
 1..522
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3180165"
 /clone_lib="NCI-CGAP_Lu24"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."
 117 a 168 c 153 g 84 t

BASE COUNT 117 a 168 c 153 g 84 t
 ORIGIN
 Query Match 33.3%; Score 515.8; DB 35; Length 522;
 Best Local Similarity 99.6%; Pred. No. 2e-110;
 Matches 517; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 878 ctgcacaaacccctgtccagtgctctgagggacagagagcggtcttacctct 937
 DB 4 CCGCACAAACCCCTGTCCAGGTGCTCTGGAGGGACAGAGAGCGGCTTACTCT 63
 QY 938 accagatgaagagaccccaacccagacctgactgagctcgccgggagacccacgc 997
 DB 64 ACCAGATGGAAGAACCCACACCCAGACCTGACTGAGCTCGGCCAGGAGACCCAGC 123
 QY 998 agcgatccacgtgtcagagagcagctgagcttactggccatggactccacatttg 1057
 DB 124 AGGCACTCCAGTGTACAGAGACAGCTGAGCTTACTGGGCCATGGACTCCACATTG 183
 QY 1058 agctctcaagatctgtgctgagagcaacaagatgtgaagattgagcgtgaggcacc 1117
 DB 184 AGCTCTCAAGATCTGTGCTGAGAGAACCAAGGATGTGAAGATTGAGCGTGGGGCACC 243
 QY 1118 tgctctcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1177
 DB 244 TGCTCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 303
 QY 1178 gccgctgcagatcaagggtggaggccgtgagtatctaccgtgtccacggtcaggctca 1237
 DB 304 GCCGCTCGAGATCAAGGGGTGGAGGCCGTGAGTATCTACCACTTCCAGGTACAGCTA 363
 QY 1238 ctgctgaggaactcagggaacacagtgaccaggaagcaggaggttgagctggggcagg 1297
 DB 364 CTGCTGAGGACTCAGGGAACAGCAGTGACCAAGAGCGAGGTGGAGCTGGGGCAGG 423

QY 1298 tgcccttctggctctctccattgccccccacgcccagatctgccccccaggaagccagaa 1357
 DB 424 TGCCCTTTTCGCTCTCTCCATTGCCCCAGGCCAGATCTGCCCCCATGAAGCCAGAA 483
 QY 1358 atgccagccgaagtgaactcctaaaggggaactccc 1396
 DB 484 ATGCCCAGCCGAAAGTGAGATCTCTAAAGGGGAAGTCC 522
 RESULT 4
 A1690580/c
 LOCUS
 DEFINITION TQ2all.x1 NCI-CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2207612 3', similar to TR:Q13193 Q13193 CBL-B TRUNCATED FORM 2.; mRNA
 ACCESSION A1690580
 VERSION A1690580
 KEYWORDS EST
 SOURCE A1690580.1 GI:4901882
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 381.
 Location/Qualifiers
 1..530
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2207612"
 /clone_lib="NCI-CGAP_Ut3"
 /tissue_type="poorly-differentiated endometrial adenocarcinoma, 2 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.45 kb. Life Technologies catalog #:
 11541-018"
 BASE COUNT 87 a 162 c 161 g 118 t 2 others
 ORIGIN
 Query Match 33.1%; Score 512; DB 12; Length 530;
 Best Local Similarity 99.6%; Pred. No. 1.5e-109;
 Matches 512; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1022 agctgcagcttactggggccatgactccacatttgagctctgcaagatctgtgctgaga 1081
 DB 514 AGCTGCAGCTTACTGGGCCATGGACTCCACATNTGASCTTGCAGATCTGTCTGAGA 455
 QY 1082 gcaacaagatgtgaagattgagcgtgaggccacctgctgctgagctgctgctgctg 1141
 DB 454 GCACAGGATGTGAGATTGAGCCGTGGGGACCTGCTCAGCTGCTGGCTG 395
 QY 1142 cctggcagactcggacagccagacctgccccctctctccgctcgagatcaaggctggg 1201
 DB 394 CTGGCAGCACTCGGACAGCCAGACCTGCCCTTCTCCGCTCGGAGATCAAGGGCTGG 335

QY	1202	aggccctgagtctaccagttccacggtccaggtctacgtctgaggaactcagggaacagca	1261
Db	334	AGGCCGTGAGTATCATTACAGTTCACAGGTACAGGTACAGGTACGTGAGGACTCAGGGAACAGCA	275
QY	1262	gtgaccagcaagcgaggagtgtagctggggcgagggtgccctttcgctccattgc	1321
Db	274	GTGACCAGGAAGCGAGGAGTTGGAGCTGGGCGAGGTGCCCTTTCGGCTCCTCCATTGC	215
QY	1322	ccccagggccagatctgccccccaggaagccagaaatgccagccgcaaaagtggagctcc	1381
Db	214	CNCCAGCGGCAGATCTGCCCCCCAGGAAGCCAGAAATGCCACGCCGAAAGTGGAGACTCC	155
QY	1382	taaaaggggaactccctccagctgcgctgggaccccccagagccctgccccggcctgaaggc	1441
Db	154	TAAAGGGGAACTCCCTCCAGCTGGCGTGGACCCAGGACCCCTGCCCGCGCTGAAGGC	95
QY	1442	caggggcaccagatgtgtctcaaggagcccccaaggcttggaagggttgtaaacc	1501
Db	94	CAGGGCACCAGATGTGTGTCTCAAGGAGAGGCCCAAGAGGCTGGAAGGGGTGTGTAACC	35
QY	1502	gaataaaactgcaagcctggtctcaaaaaa	1535
Db	34	GAATAAATGTCGACAGCTGGTCTGTCAAAAAA	1
RESULT	5		
LOCUS	BE198430	533 bp mRNA EST	26-JUN-2000
DEFINITION	ug78e05.y1 Soares mammary.gland_NMLMG Mus musculus cDNA clone IMAGE:1548512 5' similar to gb:X57110 PROTC-ONCOGENE C-CBL (HUMAN)// mRNA sequence.		
ACCESSION	BE198430		
VERSION	BE198430.1	GI:8710599	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 533)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. NCI:953860		
FEATURES	Seq primer: -40RP from Gibco High quality sequence stop: 466. Location/Qualifiers		
source	1. .533 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:1548512" /clone_lib="Soares mammary.gland_NMLMG" /sex="female (lactating)" /tissue_type="mammary gland" /lab_host="DH103" /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library is normalized. Library" was constructed by Bento Soares and M. Fatima Ronaldo.		
BASE COUNT	125 a	159 c	133 g
ORIGIN	106 t		

Best Local Similarity	85.7%	Pred No. 5e-86	0;	Gaps	0;
Matches	457;	Conservative	0;	Mismatches	76;
QY	646	gacgttcttcacccaggctcttttcagccatgcccacacacctctctcaagaactcgcagctcctg	.705		
Db	1	GACGCTCTTCACAGGCTCTTTCACGCGTGCGCCACTTTACTGAGGAATTGCCAACTCGT	60		
QY	706	gcattcaaccacccaggctacatggccttcctcaactatgatgatgaggtccaaagagcgtgt	765		
Db	61	CTGTCAACCATCTCTGGCTACATGGCTTCCTCACCTACGATGAGGTCCAAACAGCGCTG	120		
QY	766	caggctgcagggacaaagccaggcagttacattatctcttcggcccgagctgtactcgtcgggg	825		
Db	121	CAGGCTACAGGGACAAACAGGACGCTATATCTTCGGGCCAAGTGTAACCGCTGGGG	180		
QY	826	cagtgggccatcggtatgtgagctcagatggcagacatcctcagacacatccctgcgcaac	885		
Db	181	CAGTGGGGCAATGGATACGTGAGCTCGATGGAAGCATCTTCAACCATTCCTCTCAAC	240		
QY	886	aaacccctgtccagggtgctcctggaggacagaagcagcgttctacctataccacagat	945		
Db	241	AAACCTCTGCTGCAGGTCCTCTGAAGGGACAAAGAGCGGCATCTCTCTTCCTCAT	300		
QY	946	ggaaagaccccaaccacagacctgactgagctcggccaggcagaacccacgagcgatc	1005		
Db	301	GGAAAGAAACAAACCCAGACCTGACTGAGCTCTGCGGGTAGAACCTTACCAACGCATC	360		
QY	1006	cagtgctcagagagcagctgcagcttctactggccatggactccacatttgagctcgc	1065		
Db	361	CAAGTGTCAAGAGAACACTGCTCTACCAAGCCATGAATCCACGTTCCAGCTGTC	420		
QY	1066	aagatctgtctgagagcaacaaggatggtgaagatgagccgtgaggcaccctgctctgc	1125		
Db	421	AAAGATCTCGTAGAGAGGGACAAAGACGTGAGATCGAGCCCTGTGGGATCTTCTCTGC	480		
QY	1126	acgtctccttgctcgtcggcagcactcgagcagccagcactgcccctcttg	1178		
Db	481	AGCTGCTGCGTGCTGCTGCGAGATCAGACGATCAGACGACCTGTCCTCTCTG	533		
RESULT	6				
AI368245/c					
LOCUS	AI368245	398 bp	mrna	EST	15-FEB-1999
DEFINITION	QW46f09.x1 NCI-CGAP-U73 Homo sapiens cDNA Clone IMAGE:1991273 3'				
	similar to TR:Q13193 Q13193 CBU-B TRUNCATED FORM 2. ; contains				
	PTRS.13 MER22 repetitive element ; ; mRNA sequence.				
ACCESSION	AI368245				
VERSION	AI368245.1	GI:4137990			
KEYWORDS	EST				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 398)				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
	Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D.				
	Tel.: (301) 496-1550				
	Email: Robert.Strausberg@nih.gov				
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.				
	Emmert-Buck, M.D., Ph.D.				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: Greg Lennon, Ph.D.				
	DNA Sequencing by: Washington University Genome Sequencing Center				
	Clone distribution: NCI-CGAP clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LLNL at:				
	www-bio.llnl.gov/bbrp/image/image.html				
	Insert Length: 648				
	Seq primer: ~400P from Glbco				
	High quality sequence stop: 359.				
	Location/Qualifiers				
FEATURES					

Query Match

26.6%; Score 411.4; DB 34; Length 533;

FEATURES

source
 1. 398
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:191273"
 /clone_lib="NCI_CGAP-Ut3"
 /tissue_type="poorly-differentiated endometrial
 adenocarcinoma, 2 pooled tumors"
 /lab_host="DH10B"
 /notes="Organ: uterus; Vector: pCMV-SPORT6; Site:1: Salt;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.45 kb. Life technologies catalog #:
 11541-018"
 61 a 125 c 129 g 83 t

BASE COUNT 61 a 125 c 129 g 83 t
 ORIGIN

Query Match 25.6%; Score 396.4; DB 10; Length 398;
 Best Local Similarity 99.7%; Pred. No. 1.5e-82;
 Matches 397; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1132 tgcctgctcctggcagcactggagacccagacactgccccttctgcccgtcgagatc 1191
 Db 398 TGcctgctcctggcagcactggagacccagacactgccccttctgcccgtcgagatc 339
 QY 1192 aaggctggagccgtgagtatcaccagttccaggtccagctactctgaggactca 1251
 Db 338 AAGGCTGGAGGCGGTGAGTACTACCACTTCCCGGTGAGGCTACTGCTGAGGACTCA 279
 QY 1252 ggaacagcagtgaccaggaagcagggagttgagctggggcagtgccccttctggct 1311
 Db 278 GGAACAGCAGTACCAGGAAGCAGGGAGTTGGAGCTGGGGCAGGTGCCCTTTCGGCT 219
 QY 1312 cctccattgccccagggccagatctgccccccaggaagcccaagaatgccagccgaaa 1371
 Db 218 CTTCCATTGCCCGACGGCGAGATCTGCCCGCCCGCAGGAAGCCAGAAATGCCAGCGCAA 159
 QY 1372 gtgagactcctaaagggaactccctccagctgctggcctggagcccccagaccctgcccg 1431
 Db 158 GTGACAGTCTTAAGGGGACCTCCCTCCAGCTGGCTGGGCCCCAGACCTTCCCGG 99
 QY 1432 gctgaagccagggccaccagatgtcgtcctcaaggaggcccccagggtggaagggg 1491
 Db 98 GCTGAAGCCAGGGACCCAGATGTGCTCAAGGAGCCCAAGGCTGGAAGGGGG 39
 QY 1492 ttgtgaacggaataaactgcaagcctggtctgtca 1529
 Db 38 TTGTGAACGGAATAAATGCTGCAAGCCCTGGTCTGTCA 1

RESULT 7
 AA113289/c 462 bp mRNA EST 23-DEC-1997
 LOCUS
 DEFINITION
 Zm28c07.s1 stragene pancreas (#937208) Homo sapiens cDNA clone
 IMAGE:26956 3' similar to TR:G862411 G862411 CBL-B TRUNCATED FORM
 2. ; mRNA sequence.
 AA113289.1 GI:1664994
 VERSION
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 1 (bases 1 to 462)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chissoe, S., Dietrich, N., Dubucque, T., Favell, A., Gish, W., Hawkins,
 M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Wards, E., Moore,
 B., Morris, M., Parsons, J., Prange, C., Rifkin, B., Rohlfing, J.,
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.,
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 9704478
 CONTACT: Wilson RK
 MEDLINE
 COMMENT

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 WARNING: There is evidence that suggests that the 384-well parent
 plate of this clone contains both human and mouse derived clones.
 Thus, the origin of this clone is uncertain. This caution should be
 kept in mind should you use this clone.

This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Possible reversed clone: similarity on wrong strand
 Seq primer: -40M13 fwd. from Amersham
 High quality sequence stop: 367.

FEATURES

source
 1. 462
 /organism="Homo sapiens"
 /db_xref="GDB:3918385"
 /db_xref="taxon:9606"
 /clone="IMAGE:526956"
 /clone_lib="Stratagene pancreas (#937208)"
 /lab_host="SOJR cells (kanamycin resistant)"
 /notes="Organ: pancreas; Vector: pBluescript SK-; Site:1:
 EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
 Oligo dT. Pancreatic adenocarcinoma cell line. Average
 insert size: 1.0 kb. Uni-ZAP XR Vector; -5' adaptor
 sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTT 3' 99 t
 BASE COUNT 76 a 141 c 137 g 99 t
 ORIGIN

Query Match 25.4%; Score 393.6; DB 1; Length 462;
 Best Local Similarity 95.3%; Pred. No. 7.1e-82;
 Matches 443; Conservative 0; Mismatches 18; Indels 4; Gaps 4;
 QY 1066 aagatctgtctgagcaacaaggatgtgaagattgagccgtgcgggcaacctgctctgc 1125
 Db 462 AAGATCTGTGCTGAGACCAACAGGATGTGAAGATTGAGCGTGCAGGACCTTTCTGC 403
 QY 1126 agctgcctcctggctgctgctgagcactcgacag-cacagacctgccccttctgcccgtg 1184
 Db 402 AGCTGCTGCTGGTGGTGGCTGGCAGACTCGACAGCCAGACCTGCCCTTCTGCCGCTG 343
 QY 1185 cgagatcaagggtggggagccgtgagttatctaccagttccacggtcaggtactgctga 1244
 Db 342 CGAGATCAAGGGGTGGGAGGCGGTGAGTATCTACCACTTCTACGGTCAGGCTACTGCTGA 283
 QY 1245 ggaactcaggaacagcagtgaccaggaagcagggagttggaagctgggcagggtgccct 1304
 Db 282 GGATCAGGGAACAGCAGTACCAGGAAGCAGGGAGTTGGAGCCTGGCAGGTGCCCT 223
 QY 1305 ttggctcctccattgccccacggccagatctgccccccaggaagccagaaatccca 1364
 Db 222 TTNGGCTCTTCATTCGCCCCACGCGCAGATCGCCCCCAGGAAGCCCAAGAAATGCCCA 163
 QY 1365 gcgaaagttagactcctcctaaggagggaactccctccagctgagctgggacccagacc 1424
 Db 162 GCCGAAAGTGAAGTCTCTTAAAGGGGAAC-TCCNCNCTNTGCGCTGGGACCCAGGACNC 104
 QY 1425 tgccccggcctgaagccagggcaccagatgtgctgctcaaggaggccccaaggtctg 1484
 Db 103 NCINCCG-CTGAAGCCAGGGACCCACCATGTGCTGCTCAAGGGA-CCCCAAGGCTGG 46
 QY 1485 aaggggtgtgaaacccgaataaactgcaagcctggtctgtca 1529
 Db 45 AAGGGGTTGTGAACCCGAAATAAAGTGCACCAAGCTGGTCTGTCA 1

RESULT 8
 AA463722
 LOCUS

EST 24-FEB-2000

DEFINITION BP230013B10F1 Soares normalized bovine placenta Bos taurus cDNA
clone BP230013B10F1 5', mRNA sequence.
ACCESSION AW453722.1 GI:7033890
VERSION AW453722.1
KEYWORDS EST.
SOURCE Bos taurus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 440)
AUTHORS Lewin, H.A., Soares, M.B., Rebeiz, M., Pardinas, J., Liu, L. and Larson, J.H.
TITLE Bovine ESTs
JOURNAL Unpublished (2000)
COMMENT Contact: Lewin, H. A.
W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 244 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Funding for cattle EST sequencing was provided by the USDA National
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
from Washington University Genome Center. Vector Trimmer g:
Crossmatch from Washington University Genome Center PHRAP suite.
Sequences submitted are vector free and at least 200 bp in length.
PCR Primers
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTACCTCTCTAAG
Insert Length: 440 Std Error: 0.00
Plate: BP230013B10 row: F column: 1
Seq primer: AGCGGATAACAATTTCACACAGGA
High quality sequence stop: 440.
FEATURES
source
1..440
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="BP230013B10F1"
/clone_lib="Soares normalized bovine placenta"
/sex="female"
/lab_host="DH10B"
/note="Organ: Placenta; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI; The cDNA library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806."
BASE COUNT 100 a 142 c 120 g 78 t
ORIGIN
Query Match 24.0%; Score 370.8; DB 22; Length 440;
Best Local Similarity 90.4%; Pred. NO. 1.5e-76;
Matches 396; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
Qy 760 cgtctgagcctcagggcaagccaggttacattcttcggccacactgtactcgc 819
Db 1 CGTCTGACAGCTTCAGACAGACCCAGGAGGAGTACATCTCCGGCCGCTGCACCTCGC 60
Qy 820 ctggggcagtggttcagctggtatgtgctcagatggcagcatcctgcagaccatccct 879
Db 61 CTGGGGAGTGGGCGCATCGTACGCTCAGCTCAGCGGACGATCTTCAGACCATCCCT 120
Qy 880 gccacaacacccctgtccaggtgtcctctggagggagagagcgcttctacattac 939
Db 121 CTCACAAACCCCTGTTCCAGGTGCTCTCTGGAGGAGACAGAGGAGGCTTCTACCTCTAC 180
Qy 940 ccagatggaagaccacacccagactgactagctcggccaggcagaccacccagcag 999
Db 181 CCAGATGGGAAGAACCCACCCGAGCTGACCGAATCTGCCACATGGACCCGATCAG 240

Qy 1000 cgcattcacatgtcagaggagcagctgcagctctacttgggcccagaccacatttgag 1059
Db 241 CACATCCACGTCGGAGGAGCAGCTGCAGCTCTACTGGCCCATGGACATCTAGCTTGAG 300
Qy 1060 ctctgcaagatctgtctgagagcaacaagatgtgaagattgagcgtgcgggcacctg 1119
Db 301 CTCTGCAAGATCTGTGCGGAGAGCACAAGAGCTGAAGATAGAGCCCTCGGGCACCTG 360
Qy 1120 ctctgcaagctgctgcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1179
Db 361 CTCTGTAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Qy 1180 cgcctgcagatcaaggagc 1197
Db 421 CGTCTGAGATCAAGGC 438
RESULT 9
RA112513 435 bp mRNA EST 23-DEC-1997
LOCUS znm2807.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
DEFINITION IMAGE:526956 5' similar to gb:X57110 PROTO-ONCOGENE C-CBL (HUMAN);
mRNA sequence.
ACCESSION AA112513
VERSION AA112513.1 GI:1665062
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 435)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins,
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
WARNING: There is evidence that suggests that the 384-well parent
plate of this clone contains both human and mouse derived clones.
Thus, the origin of this clone is uncertain. This caution should be
kept in mind should you use this clone.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: 28M13 rev2 from Amersham
High quality sequence stop: 336.
FEATURES
Location/Qualifiers
1..435
/organism="Homo sapiens"
/db_xref="GDB:3918385"
/db_xref="taxon:9606"
/db_xref="IMAGE:526956"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTCTTTTCTTTTCTTTT 3"
BASE COUNT 94 a 136 c 121 g 82 t 2 others
ORIGIN


```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Ut3"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"

BASE COUNT      37 a      80 c      85 g      56 t
ORIGIN

Query Match      16.7%; Score 258; DB 12; Length 258;
Best Local Similarity 100.0%; Pred. No. 3.3e-50;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1271 agggcaggagtgagctggggcagggtgccccttctggctctccattgccccacagc 1330
Db 258 AAGCGAGGAGTGGAGCTGGGGCAGGTGCCCTTTTCGGCTCTCCATTGCCCCACGCG 199
QY 1331 cagatctccccaggagccagaaatgccagccgaagtgagactcctaaagggga 1390
Db 198 CAGATCTGCCCCAGGAGACCCAGAAATGCCACCGAAGTGAGACTCTTAAGGGGA 139
QY 1391 actccctccagctgctgggacccacagaccctgcccggcctgaaggccagggcacc 1450
Db 138 ACTCCCTCCAGTGGCTGGGAGCCAGCCAGACCTCCCGCCCTGAAAGCCAGGGCACC 79
QY 1451 cagatctgctcaaggagcccaaggctggaagggtgtgaaaccgaataaac 1510
Db 78 CAGATGTGCTCAAGGAGGCCCAAGGCTGGAGGGGTGTGAAACCCGAATATAC 19

QY 1511 tgccaagcctggtctgtc 1528
Db 18 TGCCAAGCCTGCTGTCTC 1

RESULT 12
AI801220/c      248 bp      mRNA      EST      14-DEC-1999
LOCUS      to86c10.x1 NCI-CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2185170 3',
DEFINITION      mRNA sequence.
ACCESSION      AI801220
VERSION      AI801220
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 248)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL      Tumor Gene Index
COMMENT      Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/obrp/image/image.html
Insert Length: 340 Std Error: 0.00
Seq primer: -400P from Gibco.

FEATURES
source
1..248
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/clone="IMAGE:2185170"
/clone_lib="NCI-CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/notes="Organ: stomach; Vector: pCMV-SPORT6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT      36 a      72 c      87 g      53 t
ORIGIN

Query Match      15.3%; Score 236; DB 13; Length 248;
Best Local Similarity 100.0%; Pred. No. 4.5e-45;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1293 gcaggtgccccttgcctccattgccccagcgccagatctgccccccaggaagcc 1352
Db 248 GCAGGTGCCCTTTCGCTCTCTCCATTGCCCCAGCCAGATCTGCCCCAGGAAGCC 189
QY 1353 cagaatgccagccgaaagtgcagactcctaaaggggaactccctccagctgcgtgg 1412
Db 188 CAGAAATGCCAGCGAAAGTGAGACTCCTAAAGGGGAACCTCCCTCCAGCTGCGTGG 129
QY 1413 acccagagaccctgcccggcctgaagccagggcaccagatgtgctcgaaggagc 1472
Db 128 ACCCCAGACCTCTCCCGCCCTGAGGCCAGCCAGATGTGCTGCTCAGGGAGC 69
QY 1473 cccaaggctgggaagggtgtgaaacccgaataaactgccaaagcctggtctgtc 1528
Db 68 CCCAAGGCTGGAAGGGGTGTGAAACCGAATAAATGCAAGCGCTGCTGTCTC 13

RESULT 13
F22931      491 bp      mRNA      EST      01-APR-1998
LOCUS      SSC17A01 Porcine small intestine cDNA library Sus scrofa cDNA clone
DEFINITION      c17a01.5', mRNA sequence.
ACCESSION      F22931
VERSION      F22931.1 GI:2578544
KEYWORDS      EST.
SOURCE      pig.
ORGANISM      Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE      1 (bases 1 to 491)
AUTHORS      Wintero, A.K., Fredholm, M. and Davies, W.
TITLE      Evaluation and characterization of a porcine small intestine cDNA
JOURNAL      Mamm. Genome 7 (7), 509-517 (1996)
MEDLINE      96327607
COMMENT      Contact: A.K. Wintero
Department of Animal Science and Animal Health, Division of Animal
Genetics, The Royal Veterinary and Agricultural University
Bulowsvej 13, 1870 Frederiksberg C, Denmark.

FEATURES
source
1..491
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="c17a01"
/clone_lib="Porcine small intestine cDNA library"
/notes="directionally cloned cDNA in Xcl-blue MRF,"

BASE COUNT      116 a      142 c      145 g      79 t      9 others
ORIGIN

Query Match      15.2%; Score 234.6; DB 37; Length 491;
Best Local Similarity 82.0%; Pred. No. 1.1e-44;
Matches 282; Conservative 0; Mismatches 60; Indels 2; Gaps 1;

QY 844 gtgagctcagatggcagcatcctgcagaccatcctgcacacacacccctgtccagtg 903
```

```

Db 1 GTGAGCTCAGATGCGAGCATCTCTCAGACCATCCTCTCACAAACCTCTGTTCACAGCA 60
QY 904 ctctctggaggagacagagagcgggtttctacctctaccagatggaagagacccacaccca 963
Db 61 CTCTCTGGAAGACAAAAGAGAGGCTTACCTCTACCTCTACCTCTGATGGGAAGAACCAACCCG 120
QY 964 gacctgactgagctcggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1023
Db 121 GACCTGACGGAGCTCTCGGTAAACGGACCCCTATCAACGCAATCCACGTGTCTGGAGGACAG 180
QY 1024 ctgacgtctactgggcatgagctccacatttgagctctgcaagatctgtgtgagagc 1083
Db 181 CTGACAGTGTACTGGGCCATGAACCTCACATACGAGCTCTGCAAGATCTGTGGCCGAGAGA 240
QY 1084 aacaagatgtagagattgagcctgctggcagcagcagcagcagcagcagcagcagcagcag 1141
Db 241 AACAAAGACGTGAAGATTGACCAATGCGGGGCACTGCTCTGCAAGCCGCTGCGCTGGGCTA 300
QY 1142 cctggcagcactcggacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1185
Db 301 CCTGGCAGAACTCAGACAGNCCAGACCTGCGCCCTTTCTGCGCCG 344

```

RESULT 14

```

AW857217 529 bp mRNA EST 19-MAY-2000
LOCUS RC2-CT0304-120200-012-e10 CT0304 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW857217
VERSION AW857217.1 GI:7952910
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-RC2-CT0304-120
200-012-e10&ts=2000-02-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 507.
Location/Qualifiers
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0304"
/dev_stage="Adult"
/Note="Organ: colon; Vector: puc18; Site: 1: Smal; Site 2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low

```

FEATURES

source

```

BASE COUNT 169 a 129 c 111 g 120 t
ORIGIN

```

```

Query Match 12.5%; Score 193.2; DB 25; Length 529;
Best Local Similarity 91.9%; Pred. No. 5.1e-35;
Matches 204; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

```

QY 923 acgggtttcttaccagatggaagagacccacacacccagacctgactgactgactgctcgcc 982
Db 308 ATGACTCTTAAGTGGGCCAGATGGAAGACCCACACCCAGACTGACTGAGCTCGGCC 367
QY 983 agcagaacccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1042
Db 369 AGGCAGAACCCACAGCAGCGCATCCACGTGTCTGAGAGGACAGCTGCTACTGGGCCA 427
QY 1043 tggactccacatttgagctctgcaagatctgtgtgagagcaacagagatgtgaagattg 1102
Db 428 TGGACTCCACATTGAGCTCTGCAAAATCTGTCTGTAGTACCAACAGGATGTGAGAGATTG 487
QY 1103 agcctgaggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1144
Db 488 AGCCGTGCGGCACCTGCTCTAGAGCTGTTCGCTAGGTGCT 529

```

RESULT 15

```

BE283475 545 bp mRNA EST 13-JUL-2000
LOCUS 601102508f1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3495025 5',
DEFINITION mRNA sequence.
ACCESSION BE283475
VERSION BE283475.1 GI:9159750
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM8544 row: p column: 02
High quality sequence stop: 542.
Location/Qualifiers
1..545
/organism="Mus musculus"
/strain="C57BL/6J (f1)"
/db_xref="taxon:10090"
/clone="IMAGE:3495025"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin"
/lab_host="DH10B"
/Note="Organ: lung; Vector: pCMV-SPORT6; Site: 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

```

BASE COUNT 97 a 188 c 167 g 93 t

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

ORIGIN

BASE COUNT

Best Local Similarity 70.2%; Pred. No. 9.4e-34;			
Matches 384; Conservative 0; Mismatches 142; Indels 21; Gaps 9;			
QY	74	cagtcaggatgctgagcctcctagagagcaatgctgaccccgctgctgcgtgagtc	133
Db	1	CGGTGAAGTGTCTCCAGCGCCCTAGAGAGCAATGAGGATCCCAAGATGTCACGGGC	60
QY	134	cccttcgctgctgggacctgctgccccgcacagcgcagctgcttcgagaggtggccatt	193
Db	61	CCCCGTCCCTGGCGGACCTGCTGCCCGCACCGCGCAGCTACTTGGAGAGGTGGCAAAG	120
QY	194	ctcggcgggcgccggcgagggcgcccggggtcccggggctctctgggactttctac	253
Db	121	CCCGCGCGAGCGCCAGGGAA---GACCCGAGGGTCCCGGTGGCGCCGATGACTTTCGG	177
QY	254	tcattctacctggccaatctggaggcccaagagcagggcaggtggc-cgcgctgctgcctcc	312
Db	178	CCATCTACCTGGCCATCTGGAGGTCAAGGCGAGCGAGCTGGCGAGCTCCTGCCACC	237
QY	313	cgggggcga-agagtgccaaagacagctcttcggggcggtccagact--cagggga	369
Db	238	CGAGGCAAAACCAAGGACGTGAACCAAGATGTTTCCGGGAGGGCTCCAGACTTCACGGGA	297
QY	370	cagctggccaa-gctggccatcatcttcagc---cacatgcacgcagagctgcagcact	425
Db	298	CAACTGGCCAAACCTGGCCCTCATCTTCAGCTCACATGCACGGCGAGCTTGAGCGCACT	357
QY	426	cttcccg-gggcaagtactgtggacacatgtaccagctcaccaggccccgcgccaca	484
Db	358	CTTCCCTGCTGGGAAGTACCGTGGCACCTGTACCAGCTACCAAGGGCTCTGCCACA	417
QY	485	ccttcggagggaagtgcggagcc-----cggtgtgctgctccctgggctgagtt	536
Db	418	TCCTCTGGCAGCGCAGAAATGTGGAGTTCGGCTGCGCTGCTTCCCTGGCGCTGCAGTTCC	477
QY	537	tgagtcctcctggtggcacctgccaccctg-tggaaacaggctgcacagccctggccttgc	595
Db	478	AGCTCCCTCCTGCTGCTGCTGCTGCCACCTGCTGGAACCAAGGCCCCACCATGCAGGCCTTGC	537
QY	596	gcaccac	602
Db	538	GGTCCAC	544

Search completed: December 7, 2000, 12:41:38
Job time: 24581 sec

Blank

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 7, 2000, 06:01:08 ; Search time 3361.48 Seconds
(without alignments)
2010.021 Million cell updates/sec

Title: US-09-434-708-1
Perfect score: 1547
Sequence: 1 cgcagcgtcccatggctct.....caaaaaaaaaaaaaaaaaa 1547

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1033670 seqs, 2183789903 residues
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl: *
1: gb_bai: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pli: *
8: gb_pl2: *
9: gb_pri: *
10: gb_pr2: *
11: gb_pr3: *
12: gb_ro: *
13: gb_sy: *
14: gb_un: *
15: em_fun: *
16: em_hum1: *
17: em_hum2: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vl: *
30: gb_htg1: *
31: gb_htg2: *
32: gb_in1: *
33: gb_in2: *
34: em_ba1: *
35: em_ba2: *
36: em_hum3: *
37: em_hum4: *
38: gb_pr4: *
39: gb_htg3: *
40: gb_htg4: *
41: gb_htg5: *
42: gb_htg6: *
43: gb_htg7: *

F22931

44: em_htg1: *
45: em_htg2: *
46: em_htg3: *
47: em_hum5: *
48: gb_pl3: *
49: gb_pr5: *
50: gb_htg8: *
51: gb_htg9: *
52: gb_htg10: *
53: gb_htg11: *
54: gb_htg12: *
55: gb_htg13: *
56: gb_htg14: *
57: gb_in3: *
58: gb_htg15: *
59: gb_htg16: *
60: gb_htg17: *
61: em_htg4: *
62: em_htg5: *
63: em_htg6: *
64: em_htg7: *
65: em_hum6: *
66: gb_htg18: *
67: gb_htg19: *
68: gb_htg20: *
69: gb_htg21: *
70: gb_htg22: *
71: gb_htg23: *
72: gb_vli: *
73: gb_vli2: *
74: gb_ba3: *
75: em_htg8: *
76: em_htg9: *
77: em_htg10: *
78: em_htg11: *
79: em_htg12: *
80: em_htg13: *
81: em_htg14: *
82: em_htg15: *
83: em_htg16: *
84: em_htg17: *
85: em_htg18: *
86: em_htg19: *
87: em_htg20: *
88: em_htg21: *
89: em_htg22: *
90: em_htg23: *
91: gb_pr6: *
92: gb_pr7: *
93: gb_sts1: *
94: gb_sts2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1539.6	99.5	1575	9	AB028645	AB028645 Homo sapi
2	1525.2	98.3	1584	11	AF117646	AF117646 Homo sapi
3	1239.2	80.1	1446	11	AF117647	AF117647 Homo sapi
4	340	22.0	3090	91	HSCCBL	X57110 Human mRNA
5	329.2	21.3	2661	57	DMU87925	U87925 Drosophila
6	327.6	21.3	2529	57	DMAJ3175	AJ223175 Drosophila
7	323.6	20.9	3241	91	HSU26711	U26711 Human cbl-2
8	323.6	20.9	3354	91	HSU26712	U26712 Human cbl-2
9	323.6	20.9	3982	91	HSU26710	U26710 Human cbl-2
10	270.6	17.3	2808	12	MMCCBL	X57111 Mouse mRNA
11	248.2	16.0	2711	72	CASNS1	J04169 Mouse Cas N
12	221.4	14.3	237653	60	AC073760	AC073760 Mus muscu

Db	532	CACACCTCTCTGGAGGGAATGTCGGAGGCCCGGTGTGTGCTGCCCTGGGCTGAGTTTGAG	591
Qy	541	tccctctgggcacactgcacccctgtggaaacaggctgcacagccctggccttgcgcacc	600
Db	592	TCCCTCTCTGGGCACCTGCCACCCCTGTGGAAACAGAGCTGCACAGCCTTGGCCTTGGCACC	651
Qy	601	aceattgaactcaacctgcaggggcaagtgtccattctgaagttcgaagcttccacagg	660
Db	652	ACCATTGACCTCACTTCAGCGGGACGTGTCCATCTTTCAGTTGCGAGCTTTCACCAGG	711
Qy	661	ctctttcagccatggcccaactcctccaagaactgcagctcctgcagttcaaacaccca	720
Db	712	CTCTTTTCAGCCATGGCCAAACACTCTCTCAAGAACTGGCAGCTCTGGACATCAACCA	771
Qy	721	ggctacatggccttctccactatgatgaaggtccaagagcgtctcgaagcctgcaggag	780
Db	772	GGCTTACATGGCCTTCTCTCACTATGATGAGGTTCCAAGAGCGTCTCAGAGCCTGCAGG	831
Qy	781	aagccaggcagttacatcttcctggcccgacagctgtactcctctggggcagtgggccatgcg	840
Db	832	AAGCCAGGCAGTTTACATCTTCGGGCCCACTGTACTCGCTGGGGCAGTGGGCCATCGC	891
Qy	841	taigtgaagctcagatggcagcactcctgcagaccatcctgtccaacaaacccctgtccag	900
Db	892	TATGTGAGCTTCAGATGGCAGATCTCTGCAGACCATCTCCCTGCCACAAACCCCTGTCC	951
Qy	901	gtgctcctggaggacagaagcagtgcttctactctctaccagatggaaagaccacaac	960
Db	952	GTGCTCTCTGGAGGCACAGAAGACGGCTTCTACCTCTTACCCAGATGGAAGACCCACA	1011
Qy	961	ccagacctgaactgagctcgcccgacagacaaacccagcagcgtccagtgctcagaggag	1020
Db	1012	CCAGACCTGACTGAGCTCGGCCAGGACGAAACCCAGCAGCGATCCACAGTGTTCAGAG	1071
Qy	1021	cagctgcagctctactggggccatggactccacatttgagctctgcgaagatctgtgctgag	1080
Db	1072	CAGCTGCAGCTCTACTTGGGCCATGGACTCCACATTTGAGCTCTGCAAGATCTGTGCTG	1131
Qy	1081	agcaacaagatgtcaagattagccgtgcgggcaactgctctcacgtcgtgcctgctct	1140
Db	1132	AGCAACAAGATGTGAAGATGTAGCCGTGTGGGGCACTCTCTCGACGTGCTGCTGGCT	1191
Qy	1141	gcttggcagcactcggacagccagaccctgccccctctgcgctcggagatcaaggctcg	1200
Db	1192	GCTTGGCAGCACTCGGACAGCCAGCACTTGCCTCTTCCCTCTCGGAGATCAAGGGCTG	1251
Qy	1201	gagccgctgagtatcaccagttccacagtcagcgtcagcgtactgctagagactcaggaacag	1260
Db	1252	GAGCCCGTGAGTATCTACCACTTTACCGGTTCAGGCTACTGTGTAGGACTCAGGGACAGC	1311
Qy	1261	agtacacgaagagcaggagttgagctggggcaggtgccccctttcggctcctccattg	1320
Db	1312	AGTGACACGAAGCAGGAGTTGAGCTTGGGAGAGTGCCTTTCCGGCTCTCTCATTTG	1371
Qy	1321	ccccacgcccagatctgccccacaggaagccagaaatgccacgcgaagtgcagctc	1380
Db	1372	CCCCACGGCCAGATCTGCCCCCCCAGAGAGCCAGAAATGCCAGCCGAAGGTGAGACTC	1431
Qy	1381	ctaaaggggaaatccccctcagctgcgtgggacccccagagacctgccccggcctgaagg	1440
Db	1432	CTAAAGGGGAATCTCCCTCTCAGCTCGCTGGGACCCAGGACCTGTCCCGGCTGAAGG	1491
Qy	1441	ccagggcacccagatgtgctgtctcaaggagcccccaaggctggaagggtgtgtcaaac	1500
Db	1492	CCAGGSCACCCAGATGTGCTGCTCAAGGAGGCCCCCAAGGGCTGGAGGGGGTGTGAAAC	1551
Qy	1501	cgaataaactgccaaagcctggctgtcaaa	1530
Db	1552	CGAAATAAACTGCCAAGCCTGTGCTGTGCA	1581

```

LOCUS       AF117647      1446 bp      mRNA                      PRI      17-JUN-1999
DEFINITION  Homo sapiens short CBL-3 protein (CBL-3) mRNA, alternatively
              spliced isoform, complete cds.
ACCESSION   AF117647
VERSION     AF117647.1      GI:4959422
KEYWORDS    human.
SOURCE      Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1446)
AUTHORS    Keane,M.M., Ettenberg,S.A., Nau,M.M., Banerjee,P., Cuello,M.,
             Penninger,J. and Lipkowitz,S.
TITLE      CBL-3: a new mammalian cbl family protein
JOURNAL    Oncogene 18 (22), 3365-3375 (1999)
MEDLINE    99289203
REFERENCE   2  (bases 1 to 1446)
AUTHORS    Keane,M.M., Ettenberg,S.A., Nau,M.M., Banerjee,P., Cuello,M.,
             Penninger,J. and Lipkowitz,S.
TITLE      Direct Submission
JOURNAL    Submitted (04-JAN-1999) Genetics Dept, Medicine Branch, DCS,
             Bethesda Naval Hospital, Bethesda, MD 20889, USA
FEATURES    Location/Qualifiers
             1..1446
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="19"
             /map="19q13.2"
             /cell_line="CF-PAC1"
             /tissue_type="pancreatic adenocarcinoma"
             1..1446
             /gene="CBL-3"
             64..1350
             /note="CBL-3"
             /gene="CBL-3"
             /note="short form, alternatively spliced; cbl family
             protein; structurally related to the Sli-1 protein,"
             encoded by the cbl gene of Caenorhabditis elegans"
             /codon_start=1
             /product="short CBL-3 protein"
             /protein_id="AAD34342.1"
             /db_xref="GI:4959423"
             /translation="MALAVAPNGRWEEARALGRAVMRLQREQCVDPLRSL
             RDLPTQALLREVAHSRRAGSGGPGGSGDFLLIYLANEAKSKROYAAK
             RSANDELPAQSLRERLOIAKLIIFSHMHAELHALPGGYKCGHMYQLTKAA
             RESCGARCVLPWAEFSLGTCHPVEPGCTALAUPTIDLCGSHVSIFEDF
             QPPELTKNHLVLAQVNHGPMFLYIDVQRLEQACRDKPGSYLYLPQKGTEN
             GQAPPOORHVSBEQQLYKWARDSTFELCKIAESNKDKVIEPCGHLGLSCC
             SDSCPTCKEPCGKEGVSIIYQFGQATDSDGNSDQEGRELELQGLVPLS
             RDPDLPKPCPNAGQKVRLLKNGSPAAALGPDPPAPA"
             276 493 443 234
             t
BASE COUNT  276 a 493 c 443 g 234 t
BRIGIN

```

	Query Match	80.1%	Score 1239.2	DB 11	Length 1446
	Best Local Similarity	90.8%	Pred. No. 2.8e-219		
	Matches 1389	Conservative 0	Mismatches 3	Indels 138	Gaps
QY	1	cgcagcgtcccatggctctgcgcgtggcccgctgggggcacacagtgggaagagccgcgc	60		
Db	52	cgcgagagctcccatggctctgcgcgtggcccgctgggggcacacagtgggaagagccgcgc	111		
QY	61	gcctctggcccgggcagtcaggatcgtctcagcgccctagaagagcaatcgctgcagcccccg	120		
Db	112	gccctctggcccgggcagtcaggatcgtctcagcgccctagaagagcaatcgctgcagcccccg	171		
QY	121	ctgtccgtgagctccctcttcgctgggggacctgtgtgcccccacacagcgacagtcgtctcga	180		
Db	172	ctgtccgtgagctccctcttcgctgggggacctgtgtgcccccacacagcgacagtcgtctcga	231		
QY	181	gagtgagcccatcttcggggggcgccggcgagcgagcgcccggggtccacggcgagctct	240		
Db	232	gagtgagcccatcttcggggggcgccggcgagcgagcgcccggggtccacggcgagctct	291		

QY 241 gggactttactatctactgccaatctgagcccaagagcagcaggtgcccgcg 300
Db 292 GGGGACTTTCACATCTACTGTCGCAATCTGGAGCCCAAGACAGCAGGTGCGCGC 351
QY 301 ctgtgctctcccccggggcgaagagtgcccaacgacagagctcttcggggggtccaga 360
Db 352 CTGTGCTCTCCCGGGGCCGAAGAGTGCACACAGCAGAGCTTTCGGGGCGGCTCCAGA 411
QY 361 ctacggcagacgtctgccaagctggccatctcttcagccacatgcacagagctcac 420
Db 412 CTCAGCCACAGCTGGCCCAAGTGGCCATCATCTTCACCCATGACGACGAGCTGCAC 471
QY 421 gcactctcccccggggcgaagtaactgtggacacatgtaccagctcaccaagccccggcc 480
Db 472 GCACCTCTTCCCGGGGGAAGTACTGTGGACACATGTACCAAGCTACCAAGCCCCCGCC 531
QY 481 cacacctcttgaggaggaagtggcgagcccggtgtgtgctgcccctgggtgagttgag 540
Db 532 CACACCTTCTGGAGGGAAGTGGCGAGCCCGGTGTGTCTGTCGCTGGCTGAGTTGAG 591
QY 541 tccccctgggacactgcacacctgtgaacacagcgtcacagccctgcttgcgcacc 600
Db 592 TCCCTCTCGGCACTGTCACACCTGTGACACAGGCTGCACAGCCCTGGCCTTGGCACC 651
QY 601 accattgacctcacctgcagcggggcagctgtccattcttgagttgcagcttaccagg 660
Db 652 ACCATTGACCTCACCTGCAGCGGACGTGTCATCTTCGAGTTCGAGCTTTCACCAAG 711
QY 661 ccttttcacatggtcccaactcctcaagaactgcagctcctgcagctcaaccaccca 720
Db 712 CTCTTTACGCCATGGCCCAACTCTCTCAAGAACTGCGACTCTCTGGCAGTCAACCAACCA 771
QY 721 ggtacatggcctctccctacatgatgaggtcccaagcgtctgcagcctgcagggac 780
Db 772 GGCTACATGGCTTCTCACCCTATGATGAGTCCAAAGACGCTCTGCAGGCCCTGCAGGAC 831
QY 781 agccaggcagttacatcttcggcccgacgtgtaactcgtcgtggcgagtgccatcggc 840
Db 832 AAGCCAGGCAG----- 842
QY 841 tatgtgagctcagatggcagcatctcgcagaccatccctgcccaaaaacccctgtccag 900
Db 843 ----- 842
QY 901 gtgctctggaggagacagaaggcagcttctaccttaccagatggaaagaccacaac 960
Db 843 -----CTACCTTACCCAGATGGAAGACCCACAC 873
QY 961 ccagactgactgactcggcagcagacagccagcagcagcagcagcagcagcagcagcag 1020
Db 874 CCAGACTGACTGAGCTCGGCCAGGAGACACCCACGAGCAGTCCACGTCAGAGGAG 933
QY 1021 cagctgcagctactggggccatggactccacatttgagctctgcaagatctgtgtgag 1080
Db 934 CAGCTGAGCTCTACCTGGGCAATGGACTCCACATTTGAGCTCTGCAAGATCTGTGCTGAG 993
QY 1081 agcaacagagatgaaagattgagcgtgaggccactgctgtgcagctgctgctggct 1140
Db 994 AGCAACAGAGATGGAATTGAGCCGTGGGCACTGCTCTGCAAGCTGCTGCTGCTGCT 1053
QY 1141 gcttggcagcactgcagacagcagcagcagcagcagcagcagcagcagcagcagcagcag 1200
Db 1054 GCCTGGCAGACTCGGACACGAGACTGCCCCCTCTGCGGCTGCGAGATCAAGGCTG 1113
QY 1201 gaggccgtgagttatctaccagttccagcgttcagcgttactctgagactcagggagcagc 1260
Db 1114 GAGCCCGTGAATCTACAGTTCTACGGTCAAGCTACTGCTGAGGACTCAGGGAACAGC 1173
QY 1261 agtaccagaaaggcagagattggagctgggcagcagctgccccttctggctctccattg 1320
Db 1174 AGTACCAGGAAGGACAGGAGTGGAGCTGGGCGAGCTGCCCCCTTTCGGCTCTCCATTG 1233
QY 1321 ccccccagcagatctgccccccaggaagcccaaatgccccagccgaagtgcagactc 1380

Db 1234 CCCCCAGGCCAGATGCCCCCAGGAAGCCAGAAATGCCACGCCGAAGATGAGCTC 1293
QY 1381 ctaaaaggaactccctccagctgcctgggacccagacccttgcctggcctgaagg 1440
Db 1294 CTAAGGGGAATCCCTCCAGCTGCGCTGGACCCAGGACCCCTGCCCGGCTGAAG 1353
QY 1441 ccagggcacccagatgtctgtcctcaaggagcccccaaggcgtggaagggtgtgaaac 1500
Db 1354 CCAGGCCACCCAGATGCTGCTCAAGGAGGCCCAAGGGCTGGAAGGGGTGTGAAAC 1413
QY 1501 cgaataaactgccaagcctgctgtctaa 1530
Db 1414 CGAAATAAAGTCCAAAGCCCTGCTGTCCA 1443
RESULT 4
HSCBEL HSCBEL 3090 bp mRNA PRI 12-DEC-1992
LOCUS Human mRNA for c-bcl proto-oncogene.
ACCESSION X57110
VERSION X57110.1 GI:29730
KEYWORDS cbl oncogene; nuclear protein; oncogene cellular.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3090)
AUTHORS Blake, T.J., Shapiro, M., Morse, H.C. III. and Langdon, W.Y.
TITLE Direct Submission
JOURNAL Submitted (02-JAN-1991) W.Y. Langdon, INSTITUTE OF MEDICAL &
VETERINARY SCIENCE, IMVS, DIVISION OF HUMAN IMMUNOLOGY, BOX 14
RUNDEL WALL POST OFFICE, ADELAIDE SA 5000, AUSTRALIA
REFERENCE 2 (bases 1 to 3090)
AUTHORS Blake, T.J., Shapiro, M., Morse, H.C. III. and Langdon, W.Y.
TITLE Direct Submission
JOURNAL Submitted (01-JAN-1991) Langdon W.Y., Institute of Medical und
Veterinary, Science, IMVS, Division of Human Immunology, Box 14
Rundel Wall Post Office, Adelaide SA 5000, Australia
REFERENCE 3 (bases 1 to 3090)
AUTHORS Blake, T.J., Shapiro, M., Morse, H.C. III. and Langdon, W.Y.
TITLE The sequences of the human and mouse c-bcl proto-oncogenes show
a rearrangement of the c-bcl proto-oncogene in HUT78 T-lymphoma
cells results in a truncated protein
JOURNAL Oncogene 7 (4), 757-762 (1992)
MEDLINE 92228506
FEATURES Location/Qualifiers
source 1..3090
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3090
/gene="c-bcl"
/evidence=experimental
1..3090
/gene="c-bcl"
149..2869
/gene="c-bcl"
/note="c-bcl protein"
/codon_start=1
/protein_id="CAA40393.1"
/db_xref="GI:29731"
/db_xref="SWISS-PROT:P22681"
/translation="MAGNVRKSGAGGTGGSGGSLGLMKDAFQPHHHHHHLS
PHPTGYDKMKVCKLMDKVKLQNLKALNSPYILDLPLDTYQHLTILSR
EGKMTYGENEYFVFMKLMKTKQTSIFKEGKRYEENSQPRNLKLSLIFSH
MLAEKGFPSGLFGQDTFRITKADAEFWKAFGEKTIYPKSFQALHEVHPISG
LEMAKSTIDLTCTNDYISVFEDIFRLEFQPSWSSLRNWNSLAVTHPGYMAETLYDE


```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3"
/dev_stage="embryonic"
/map="660"
/clone="pBluescript-1255"
/gene
328..1674
CDS
328..1674
/gene="cbl"
/gene="cbl"
/codon_start=1
/protein_id="CA11149.1"
/db_xref="GI:2739273"
/translation="MATRGSRTVRSQPKIPFSLSLKHLGAISEACVSRSLSTDKKTL
EXTKMLDVKVLCQOPKRNKLNKSPFILDILDIYORLRLIYKSNEDQMLHLHANEH
ENFVILNMRKCKQAIKFKEGKMFEDNSHYRNLTKLSLVSFSLSELKAIKPNF
VFAGDQFRITKADAADFWKSNFNSLVPKIFROELNKHVPIISLEAMAKTIDIL
TNDFTSNREDFWFTLFOFWNTLERNWLAVTPGVVAFILYDVKARLQRYILKA
GSTVFRSLTRLCQWALGVYTAEGEILQIPONKSLCOALLDGHREGFVLYPDQAYN
PLUSSAVQSTEDHITVITQELICEMSTFQCKIKCAENKDKRIEFCGHLCTPC
LTSQVDSQSGQCPFCRAEIKGTEIVDAFDFRKHNRNVTNGRQQQEEEDIEV"
BASE COUNT 657 a 561 c 564 g 647 t
ORIGIN

```

```

Query Match 21.2% Score 327.6; DB 57; Length 2529;
Best Local Similarity 51.6%; Pred. No. 3.9e-51;
Matches 541; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

```

```

QY 328 gccaacagcagctctccggcggggtccagactccaggggagcagctggcgaagctggcc 387
DB 697 GGCAGGAGAGATGTTCCGACGAGACTCCCACTACCGCGCAATCTCACCAAGCTCAGC 756
QY 388 atcatcttgcacacatgcacgcagctgcacacactcttccccgggggaaagtactgt 447
DB 757 CTGGTCTTCTCCACATGCTCAGCGAACTGAAGGCCATAITCCCCAACGGTGTCTTTGCC 816
QY 448 gacacatgtaccagctcaccagggcccgcccccacacctcttgaggagaaagtggga 507
DB 817 GGGGATCAATTCGGATACCAAGAGCGGATCGCGCTGACITTTGGAGAGAGCAATTCGGT 876
QY 508 gccgggtgtgtgcctgggtgtgagttgagttccctcctgggcaactggcaccctgtg 567
DB 877 AACAGCATATGGTTCCCTGGAAGATCTCCGGCAGGAGCTTAACAAAGTACATCCCAT 936
QY 568 gaaccagcgtgcacagcttgccttgcgacacacacacacacacacacacacacacac 627
DB 937 ATCTCCGGCCCTGGAGCCATGGCCCTAAAGACCACTATCGATCTTTACCTGCAAGACTTC 996
QY 628 gtgtccatcttcgagttcgagcttccacaggtctcttcagccatggccacacactctc 687
DB 997 ATTTCACACTTCGAGTTCGACGTCTTACACGCTCTTCCAGCCTTGGGTGACCTGTCTA 1056
QY 688 aagaactggcagctcctggcagtcacacacacacacacacacacacacacacacacac 747
DB 1057 CGCAACTGGCAGATCTGGCGCTCACATCCGGGCTACGTGCGGTGTTCTCACATACGAC 1116
QY 748 gaggtcccaagagcgtctgcagcctgcagggagacacacacacacacacacacacac 807
DB 1117 GAGGTGAAGTTCGGCTACAGCGCTACATCTCAAGGGGGCAGCTACGTTTCCGGCTC 1176
QY 808 agctgtactcgcctggggcagtgggccatcggtctatgtgagtcagatggcagacactg 867
DB 1177 TCCTGCACGCGATTCGGCCCAATGGGCCATGGCTACGTACCTGCGGAGGAGAGATTCTG 1236
QY 868 cagaccatcctggcacaacacacacacacacacacacacacacacacacacacacac 927
DB 1237 CAGACAAATCCCTCAGAACAAAGTCGCTGTGCGGCGCTGTCTGATGGCCATCGCAGAGGC 1296
QY 928 ttctacctctaccagatggaaagaccacacacacacacacacacacacacacacacac 987
DB 1297 TTCTACTTTGACCCAGATGGCCAGGATACATCCGGATCTGTCTGCTCTCCCTTCAAGT 1356

```

```

QY 568 gaaccagcgtgcacagcctgccttgcgacacacacacacacacacacacacacacac 627
DB 923 ATCTCGGCGCTGGAGCCATGGCCCTTAAGACCACTATCGATCTTACTGTGACAGACTTC 982
QY 628 gtgtccatcttcgagttgagcttcttccacagcgtcttccagccatggcgaacacactc 687
DB 983 ATTTCACACTTCGAGTTCGACGTCTTACACGCGCTTTCACAGCCCTGGGTGACACTGCTA 1042
QY 688 aagaactggcagcctcctggcagtcacacacacacacacacacacacacacacacacac 747
DB 1043 CGCAACTGGCAGATCTGGCCCTCACATCCGGGCTAGTGGCTTCTCTACATACGAC 1102
QY 748 gaggtcccaagagcgtctgcagcctgcagggagacacacacacacacacacacacacac 807
DB 1103 GAGGTGAAGGCTCGCTACAGCGCTACATCCCTCAAGGGGGCAGCTACGTTTCCGGCTC 1162
QY 808 agctgtactcgcctggggcagtgggccatcgtgctatgtgagctcagatggcagcactcg 867
DB 1163 TCCTGCACGCGATTCGGCCCAATGGGCCATCGCTACGTAACTGCCGAGGGAGAGATTCTG 1222
QY 868 cagaccatccctgcacacacacacacacacacacacacacacacacacacacacacac 927
DB 1223 CAGACAAATCCCTCAGAACAAAGTCGTGTGCGAGGGCTCTCTGATGGCCATCGAGGGC 1282
QY 928 ttctacctctaccagatggaaagaccacacacacacacacacacacacacacacacacac 987
DB 1283 TTCTACTTTGACCCAGATGGCCAGGCTACATCCGGATCTGTCTGCTCGCTTCAAGT 1342
QY 988 gaaccacagcagcgtcaccagctgcagagagcagctcagctctactctggcgaagcag 1047
DB 1343 CCACAGAGGACACATAACCGTTTCCCAAGAGCAATACGAACTATACGTGAATGGC 1402
QY 1048 tcacattgtgctgcagatctgtgtgagagcaacacacacacacacacacacacacacac 1107
DB 1403 AGCACCTTTGAGTGTGCAATTTTGGGGAGAGACGACAAAGATATCGGCATCGAGCC 1462
QY 1108 tggggcacctgtcagcgtcctcctggctggctggcga--gcactcgacacagcag 1164
DB 1463 TGTGGCCACTGTGTGCACTCCCTGCTTACCTCTGCAAGTGTGATTCGAGGGACAG 1522
QY 1165 acctgccccttgcgcgtgcgagatcaagggtggga 1202
DB 1523 GCGTGGCCCTTGTGCGGGCCGAATCAAGGGCACCGA 1560

RESULT 6
LOCUS Drosophila melanogaster mRNA for cbl gene. INV 28-DEC-1997
DEFINITION Drosophila melanogaster mRNA for cbl gene.
ACCESSION AJ223175
VERSION AJ223175.1 GI:2739272
KEYWORDS cbl gene; proto-oncogene.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
Drosophilidae; Drosophila.
1 (bases 1 to 2529)
REFERENCE Hime,G.R.
AUTHORS Direct Submission
TITLE Submitted (22-DEC-1997) Hime G.R., Trescowthick Research
JOURNAL Laboratories, Peter MacCallum Cancer Institute, Locked Bag 1, A
Beckett St., Melbourne VIC, 3000 AUSTRALIA
2 (bases 1 to 2529)
REFERENCE Hime,G.R., Dhungat,M.P., Ng,A. and Bortell,D.D.
AUTHORS D-Chi, the Drosophila homologue of the c-Cbl proto-oncogene,
TITLE interacts with the Drosophila EGF receptor in vivo, despite lacking
JOURNAL C-terminal adaptor binding sites
MEDLINE Oncogene 14 (22), 2709-2719 (1997)
FEATURES 9732115 Location/Qualifiers
source 1..2529

```


REFERENCE	2 (bases 1 to 2808)
AUTHORS	Langdon,W.
TITLE	Direct Submission
JOURNAL	Submitted (01-JAN-1991) Langdon W.Y., Institute of Medical and Veterinary, Science, IMVS, Division of Human Immunology, Box 14 Rundel Mail Post Office, Adelaide SA 5000, Australia
REFERENCE	3 (bases 1 to 2088)
AUTHORS	Blake,R.J., Shapiro,M., Morse,H.C. III. and Langdon,W.Y.
TITLE	The sequences of the human and mouse c-bcl proto-oncogenes show v-cbl was generated by a large truncation encompassing a proline-rich domain and a leucine zipper-like motif
JOURNAL	Oncogene 6 (4), 653-657 (1991)
MEDLINE	91232862
REFERENCE	4 (bases 1 to 2808)
AUTHORS	Blake,R.J. and Langdon,W.Y.
TITLE	A rearrangement of the c-bcl proto-oncogene in HUT78 T-lymphoma cells results in a truncated protein
JOURNAL	Oncogene 7 (4), 757-762 (1992)
MEDLINE	92238506
FEATURES	<p>Location/Qualifiers</p> <p>1..2808</p> <p>/organism="Mus musculus"</p> <p>/db_xref="taxon:10090"</p> <p>1..2808</p> <p>/gene="c-cbl"</p> <p>/evidence=experimental</p> <p>1..2808</p> <p>/gene="c-cbl"</p> <p>97..2787</p> <p>/gene="c-cbl"</p> <p>/codon_start=1</p> <p>/product="c-cbl protein"</p> <p>/protein_id="CAA40394.1"</p> <p>/db_xref="GI:50315"</p> <p>/db_xref="SWISS-PROT:P22892"</p> <p>/translation="WAGNVKSSGAGGGGSGAGGLGLMKDAFQPHHHHLSPH PCTYDKMVECKWMLKVRQLQNPVALNKNPPYILDLPDYQHLRTVLSRYG KMETIGENYFVFVFNLMKKTKQTLSLKEGKERYEENSQPRNLTSLISFHHM AILGTFPGLQGGTITFKADAEFWKATGKTIWPKSFROALHEVHPISSGLE AVALKSTDLTCDNYSVFEDITFLPQWSSLNWNLSAVTHPGVNAFLTYDEVK ARLOFIHKPGSVYFRLSCTRIGOWAIGYTAGDGNLTQIPIHKNKLFQALIDGFRFG YLFPQGNRNPLTIGCEPTPOQHIKVOICAEKNDKXIEFGCHLMKCSLITSQWES EGCGPFCFCEIKTEPIVYDPRGSGSLRFGAEPSPNYDDDDDERADDSLFLM MKELAGAKVERSPFSFMAQPSLPVPRDLQORAPVPASTSVLGTAKAAGSLH HKDKPLPIPTLDLVPFPPOPQPSVGAETPQRRLPCTGDCPSRDKLPFPVSSR DTSWLSRPIKPVPAITPNPGDPWNGRELTNHRSLPFLSPQMEPRADYPRGLGSTEYLM DTSMTNLSVPAPESEHRIKIPSSANALYSLAARPLMPKLPPCEGCESEDTEYML TTSRPYQVQPEKPEPLAATOSSRACDDQIDTSCTEAMYTIIQOALSVAENASG ENLATAHSTGTPESEEDSDGIDVPEKPPVAVLARITSDISNASSFGWLSLGDGP TFNFGESQVPEPPKPPFRRINSEKASSQCGGATNPVATPAPSPQUSSEIERLMS QGYSYDQIKALVIAHNHMEAKNILRFVVS-SSPAHAT"</p>
BASE COUNT	691 a 796 c 693 g 628 t
ORIGIN	
Query Match	17.5%; Score 270.6; DB 12; Length 2808;
Best Local Similarity	57.8%; Pred. No. 1.2e-40;
Matches 535; Conservative	0; Mismatches 339; Indels 51; Gaps 1;
QY	303 gctgcctcccggggcgaaggagtgccaacgacgagctcttcggcgcggtccagaact 362
Db	471 GCAGACTATCAGCCTCTTCAAGGAGGGAAGAAGAGATGTATCAGGAGAAATCCCAGCC 530
QY	363 caggcgacagtgcgcacagtgcgcattatcttcagccacatgcacgcagctgcacgc 422
Db	531 TAGCGCAAACTGACCAAAATATCCCTGATCTTCAGCCACATGCTGCGCAACTGAAGG 590
QY	423 actcttcocggggcaagttactgtgacacatatcagctcaccagcccccgcacca 482
Db	591 CATCTTTCCGAGCGGACTCTTCCAAGGAGACACTTCCGGATTACTAAGACTGATGTCG 650
QY	483 caccctcttgaggaaagtgtcgagccggtgtgtgctgcctcgggctgagttgagtc 542


```
* 102077 105527: contig of 3451 bp in length
* 105528 105627: gap of unknown length
* 105628 110955: contig of 5328 bp in length
* 110956 111055: gap of unknown length
* 111056 172992: contig of 61937 bp in length
* 172993 173092: gap of unknown length
* 173093 190577: contig of 17485 bp in length
* 190578 190578: gap of unknown length
* 190678 196355: contig of 5678 bp in length
* 196356 196455: gap of unknown length
* 196456 210920: contig of 14465 bp in length
* 210921 211020: gap of unknown length
* 211021 215114: contig of 4094 bp in length
* 215115 234194: contig of 18980 bp in length
* 234195 234294: gap of unknown length
* 234295 237653: contig of 3359 bp in length.
FEATURES
    Location/Qualifiers
        source
            1..237653
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone_lib="RP23-33H23"
            /clone_lib="RPCI mouse BAC library 23"
BASE COUNT 58023 a 59212 c 60636 g 58282 t 1500 others
ORIGIN
Query Match 14.3%; Score 221.4; DB 60; Length 237653;
Best Local Similarity 75.5%; Pred. No. 3.6e-32;
Matches 289; Conservative 0; Mismatches 91; Indels 3; Gaps 1;
QY 1 cgcagctccatgctgctggcgggtgcccgtggggcgagcagtggaagagggccgc 60
Db 43822 CCCAGGGACCCATGCTGCGGCGGAGCCCCGCGAGGGTGGCAGCGGGCGAGCCACC 43881
QY 61 gccctggcggcgagcagtgctgcagcgctcgaagcgaagcgaatgcgtcgacccccc 120
Db 43882 GCCTTAGCGCGCGGTGAAGTGTCTCCAGCGCTAGAGAGCAATGCAGGGATCCAGG 43941
QY 121 ctctcgtgagtcctccttgcgtggggagcctgtcccccgcagcagctgttcga 180
Db 43942 ATGCTACGGGGGCCCCGCTGCTGCGGGAGCTGCTGCCCGGACCGCGAGCTACTTGA 44001
QY 181 gagggtgcccattctcggcgggcgccggcgaggcgcccggggtcccgcggtct 240
Db 44002 GAGTGGCAAGCCCGCGCGAGGCGCAGGGAA--GACCCGAGGGTCCCGGTGCGCC 44058
QY 241 ggggaactttactatctacctggccaatcttgagggcccaagagcgagtgggcgcg 300
Db 44059 GATGACTTTCTGGCCATCTACCTGGCCAACTCTGGAGGTCAAAGCGCAGGAGTGGCGGAG 44118
QY 301 ctctgctcccggggcccgaagagtgcccaacgacgagctcttcggggcggtccaga 360
Db 44119 CTCCTGCCACCCGAGCAAAAGGAGCTGAACACGAGTGTTCGCGGAGGCTCCAGA 44178
QY 361 ctgaggcgacgctggccagct 383
Db 44179 TTCAGGTGAGCGCTAACCTACCT 44201
RESULT 13
CEDNASL11 2230 bp DNA INV 15-NOV-1995
LOCUS
DEFINITION C.elegans DNA for sli-1 gene.
ACCESSION X89223
VERSION X89223.1 GI:895766
KEYWORDS negative regulator; sli-1 gene.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 2230).
AUTHORS Yoon,C.H., Lee,J., Jongeward,G.D. and Sternberg,P.W.
TITLE Similarity of sli-1, a regulator of vulval development in C. elegans, to the mammalian proto-oncogene c-cbl Science 269 (5227), 1102-1105 (1995) 95381043
MEDLINE
REFERENCE 2 (bases 1 to 2230) Yoon,C.H.
AUTHORS
Direct Submission
TITLE Submitted (28-JUN-1995) C.H. Yoon, CALIFORNIA INSTITUTE OF TECHNOLOGY, Div. of Biology, Pasadena, CA91125, USA
JOURNAL
FEATURES
    Location/Qualifiers
        source
            1..2230
            /organism="Caenorhabditis elegans"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone_lib="Barstead"
            /map="-20 to -24.50 map units"
            65..1813
            /gene="sli-1"
            65..1813
            /gene="sli-1"
            /function="negative regulator of LET-23-mediated signaling"
            /codon_start=1
            /protein_id="CAA61507.1"
            /db_xref="GI:895767"
            /db_xref="SPTREMBL:Q19019"
            /translation="MGSINTIPRIHRTVNGTGNARFVPTNNSTEALTLSPRAVPS TVSLFIPSAEMPGTCSDEDRFLKACKMDOVKVSKSPRLNKNPPFLDILP DTVHMLFTQNDILONDYKILESMINKCKEIKLFKTSALYNQSEERRKLT KMSLTFSHMLFEIKALFFPIEDREFRMKRAESFWSHHFKKIVPWSFFTALE KHGISTGKMAAELKATIDLSGDDFISNFEFDTFLTFPFLKILNWTULTIAPG YCALFYDEVKKEKLEKLPKPSYIFLSCITREGOWAIGYVADPGKIYTIPOKSLI QALHEGHKEGYIYPNGRQDINLSKIMDVQADRVQVISEQVELYCEMGTTPELCKI CDNEKNKIETPCGHLKCAKLANWQDSGGNTCPFCRVEIKGTNRVITDRFPYV EIEKANAARAEKILSLVDPVPTVVSQCSOLLDASNSIPSDVDELPLVPPPLPP KALGTLDTLNSSQTSISYNNIKELNVEISGEALQVWNRQAFSIQALPLPRLSAS EHOPHEPDYTNISERE"
BASE COUNT 701 a 487 c 429 g 613 t
ORIGIN
Query Match 12.9%; Score 199.4; DB 33; Length 2230;
Best Local Similarity 53.5%; Pred. No. 1.8e-27;
Matches 466; Conservative 0; Mismatches 396; Indels 9; Gaps 2;
QY 365 ggcgacagctggcgaagctggccatctcttcagccatcagcagcagcagcagc 424
Db 531 GACGGAAGCTTACGAAATGTCACTACATTTTCACATATGCTTTTCGAGATTAAAGCAT 590
QY 425 tcttcccgggggcaagtgactgtggacacatgtaccagctcaccgaagcccccacaca 484
Db 591 TATTTCCGGAAGGTATCTATATTGAAGACCGGTTTCGGATGACAAAGAAAGGCCGAAA 650
QY 485 cctctgaggggaaggttgcgagcccggtgtgtgctgctcctgggctgagttgagtc 544
Db 651 GCTTTTGAGTCATCATTTTACAAAAAACAATGTACCTTGGTCAACATTTTACTG 710
QY 545 tcttgggcaactggca---cctgtggaaaccagggtgcacagccctggccttgccaca 601
Db 711 CATTAGAAAAACCACTGATGATCAACGATAGGAAAAATGGAAGCAGCCGGAATATAAGGTA 770
QY 602 ccatgacctcaactgcagcggggcagctgtccatcttcgagttgacgtcttcaccaggc 661
Db 771 CGATAGACTTGAGCGAGATGATTTTATTCGAATTTTGAGTTTGATGTGTTTACAGGT 830
QY 662 tctttcagccatggcccaactcctccaagaactggcagctcctggcagtcacaccaccag 721
Db 831 TATTTACCTTTCAAAACACATGATCAAAAATTTGCAAAACACTCACCACCGCCATCCG 890
QY 722 gctacatgctccttcacactatgatgaggtcccaagcagctctgcagcctgcagggaca 781
Db 891 GATACTGTCTATTTCTCATATACGATAGGTCAAAAAACGGTTAGAAAAAATTACGAAA 950
```

QY 782 agccaggcaggtatcatcttcgcccagcagctgtactgcctgggagctgggcccagctggct 841
 DB 951 RACCTGGAGCTACATCTCCGGTTATCATGCACAGCTCTGGACAATGGGCAATAGGAT 1010
 QY 842 atgtgagctcagatggcagcagctctgcagacattccctgccacacacccctgtccag 901
 DB 1011 ACGTAGCTCCGGTGAAGAATTTATCAGACAATACACAGATAAAAGTTTGATTCAAG 1070
 QY 902 tgcctctgagggacagagagcagctcttacccttaccagatggagaaagaccacacacc 961
 DB 1071 CACTACATGAGGGCCATAAAGAGATTTTATATTTACCGAACGGTAGAGATCAAGATA 1130
 QY 962 cagacctgactgagctcgccagcagagacacccagcagcagctccagctgcagagagc 1021
 DB 1131 TTAATTATCAATTTGATGATGTGCCACAGCGGACAGTCAAGTACACAGTGAAC 1190
 QY 1022 agctgagctctactgggcccagtgactccacattgtagctctctgaagatctgtgctgaga 1081
 DB 1191 AATACGAGTTCTATGTGAGATGGGCACACATTCGAGTTGTGCAAAATTTGTGACGATA 1250
 QY 1082 gcaacaagatgtgaagattgacggtgagcgggacacctgctcagagctgctgctggctg 1141
 DB 1251 ACGAGAGAACATCAAAATGAGCCATGTGACATTTGCTCTGCGCAAAATGTTGCTA 1310
 QY 1142 cctggcagcactcgga-----cagccagacctgcccctcttgcgctgcagatcaag 1195
 DB 1311 ACTGGCAGGATTCGATGCTGGTGGTGGCAACACATGTCCTTCTGCGGCTACGAATCAAG 1370
 QY 1196 gctgggagccgtgagtatctaccagttcca 1226
 DB 1371 GAACAAATCGTGATTTATTCACAGGTTCAA 1401

RESULT 14
 AC014994/c
 LOCUS AC014994 36890 bp DNA HTG 16-NOV-1999
 DEFINITION Drosophila melanogaster, ** SEQUENCING IN PROGRESS **, in ordered pieces.
 ACCESSION AC014994
 VERSION AC014994.1 GI:6436341
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 36890)
 AUTHORS Adams, M. and Venter, J. C.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
 COMMENT This sequence was identified as CDL:10212090 by the submitter.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source Location/Qualifiers
 1..36890
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 11572 a 7373 c 7329 g 10616 t
 ORIGIN
 Query Match 11.6%; Score 179.2; DB 41; Length 36890;
 Best Local Similarity 65.5%; Pred. No. 4e-24;
 Matches 262; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 518 tgcctgcccgtggctgagttgagctcctctgggacccctgccacccctgggaaccagct 577
 DB 30388 TGGTTCCTCGGAAATCTCCGCGCAGGAGCTTACAAAGTACATCCATTAATCTCCGGCC 30329

QY 578 gcacagccctggccttgcgcacacacattgacctcaactgcagcgggacagctgtccatct 637
 DB 30328 TGGAGCCCATGGCCCTAAAGACCACTATCGATCTTACCTGCAACGACTTCTTCCAACT 30269
 QY 638 tgcagttcagctcttcaccagctctcttcagcctatggccacacactcctcaagaactgc 597
 DB 30268 TCGAGTTCCAGCTCTTCACACCCCTCTCCAGCCCTTGGTGACCTGCTACGCAACTGGC 30209
 QY 698 agctcctggcagctcaacacacccagctacatggccttccctcaactatgatgaggtcccaag 757
 DB 30208 AGATTTCGCCCTCACACATCCGGCTACGTGGCGTTTCTTCACATACGACGAGTGAAGG 30149
 QY 758 agcgtctcagccttcagggagacagcagcagcttaccatcttccggccagcagctgaactc 817
 DB 30148 CTCGGCTACAGCGCTACATCTCAAGCGCGGACGCTACGTTTTCGGGCTCTCTCGACGC 30089
 QY 818 gctcgtggcagctgggcccctcgtgctatgtgagctcagatggcagcactcctgcagaccatcc 877
 DB 30088 GATTGGCCCAATGGCCATCGCTACGTACGTAACTGCGGAGGAGAGATCTCGCAGACATCC 30029
 QY 878 ctgcacacacacccctgtccaggtgctcctcgtggagggaca 917
 DB 30028 CTCAGAACAAATCGCTGTGCCAGCGCTGCTCGATGGCCA 29989

RESULT 15
 AC010033
 LOCUS AC010033 129226 bp DNA HTG 20-MAY-2000
 DEFINITION Drosophila melanogaster clone RPC198-315, ** SEQUENCING IN PROGRESS **, 75 unordered pieces.
 ACCESSION AC010033 AC023749
 VERSION AC010033.7 GI:7958624
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 129226)
 AUTHORS Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K., Bodota, B., Burrows, J., Bowie, S., Brooks, A., Buhay, C., Buncick, C., Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, E., Cox, C., David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N., Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L., Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M., Hollway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondejewski, N., Kong, X., Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozado, R.J., Lu, J., Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R., Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wabnah, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.

TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 129226)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (11-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 COMMENT On May 20, 2000 this sequence version replaced gi:5996835 gi:5996719.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

```
----- Project Information
Center project name: DRCC
Center clone name: RPC198-315
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 62108 bases at least Q40
Consensus quality: 84593 bases at least Q30
Consensus quality: 95829 bases at least Q20
Estimated insert size: 100104; sum-of-contigs estimation
Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* As soon as it is available and the accession number will
* be preserved.
*
1 8568: contig of 8568 bp in length
8669 8668: gap of unknown length
8669 13396: contig of 4728 bp in length
13397 13496: gap of unknown length
13497 17696: contig of 4200 bp in length
17697 17996: gap of unknown length
17997 21860: contig of 4064 bp in length
21861 21960: gap of unknown length
21961 25098: contig of 3138 bp in length
25099 25198: gap of unknown length
25199 27960: contig of 2762 bp in length
27961 28060: gap of unknown length
28061 30796: contig of 2736 bp in length
30797 30896: gap of unknown length
30897 33631: contig of 2735 bp in length
33632 33731: gap of unknown length
33732 36353: contig of 2622 bp in length
36353 36453: gap of unknown length
36454 39018: contig of 2565 bp in length
39019 39118: gap of unknown length
39119 41507: contig of 2389 bp in length
41508 41607: gap of unknown length
41608 43946: contig of 2339 bp in length
43947 44046: gap of unknown length
44047 46295: contig of 2249 bp in length
46296 46395: gap of unknown length
46396 48593: contig of 2198 bp in length
48594 48693: gap of unknown length
48694 50813: contig of 2120 bp in length
50814 50913: gap of unknown length
50914 53017: contig of 2104 bp in length
53018 53117: gap of unknown length
53118 55095: contig of 1978 bp in length
55096 55195: gap of unknown length
55196 57155: contig of 1960 bp in length
57156 57255: gap of unknown length
57256 59209: contig of 1954 bp in length
59210 59309: gap of unknown length
61213 61313: contig of 1904 bp in length
61314 63167: gap of unknown length
63168 63267: gap of unknown length
63268 65097: contig of 1830 bp in length
65098 65197: gap of unknown length
65198 66930: contig of 1733 bp in length
67030 67030: gap of unknown length
68684 68784: contig of 1654 bp in length
68785 70393: contig of 1609 bp in length
70394 70493: gap of unknown length
72081 72081: contig of 1588 bp in length
72082 72181: gap of unknown length
72182 73749: contig of 1568 bp in length
73750 73849: gap of unknown length
73850 75387: contig of 1538 bp in length
75388 75487: gap of unknown length
75488 77000: contig of 1513 bp in length
77001 77100: gap of unknown length
77101 78600: contig of 1500 bp in length
78601 78700: gap of unknown length
78701 80199: contig of 1499 bp in length
80200 80299: gap of unknown length
80300 81739: contig of 1440 bp in length
81740 81839: gap of unknown length
81840 83268: contig of 1429 bp in length
83269 83368: gap of unknown length
83369 84792: contig of 1424 bp in length
84793 84892: gap of unknown length
84893 86250: contig of 1358 bp in length
86251 86350: gap of unknown length
86351 87676: contig of 1326 bp in length
87677 89094: contig of 1318 bp in length
89095 89194: gap of unknown length
89195 90449: contig of 1255 bp in length
90450 90549: gap of unknown length
90550 91801: contig of 1252 bp in length
91802 93127: gap of unknown length
93128 93227: gap of unknown length
93228 94440: contig of 1213 bp in length
94441 94540: gap of unknown length
94541 95730: contig of 1190 bp in length
95731 95830: gap of unknown length
95831 97015: contig of 1185 bp in length
97016 97115: gap of unknown length
97116 98210: contig of 1095 bp in length
98211 98310: gap of unknown length
98311 99370: contig of 1060 bp in length
99371 99470: gap of unknown length
99471 100529: contig of 1059 bp in length
100530 100629: gap of unknown length
100630 101680: contig of 1051 bp in length
101681 101780: gap of unknown length
101781 102823: contig of 1045 bp in length
102824 102923: gap of unknown length
102924 103948: contig of 1023 bp in length
103949 104048: gap of unknown length
104049 105054: contig of 1006 bp in length
105055 105154: gap of unknown length
105155 106133: contig of 999 bp in length
106134 106233: gap of unknown length
106234 107244: contig of 991 bp in length
107245 107344: gap of unknown length
107345 108328: contig of 984 bp in length
108329 108428: gap of unknown length
108429 109408: contig of 980 bp in length
109409 109508: gap of unknown length
109509 110479: contig of 971 bp in length
110480 110579: gap of unknown length
110580 111536: contig of 957 bp in length
111537 111636: gap of unknown length
111637 112557: contig of 921 bp in length
112558 112657: gap of unknown length
112659 113571: contig of 914 bp in length
113572 113671: gap of unknown length
113672 114559: contig of 888 bp in length
114560 114659: gap of unknown length
114660 115542: contig of 883 bp in length
115543 115642: gap of unknown length
115643 116515: contig of 874 bp in length
116516 116616: gap of unknown length
116617 117461: contig of 845 bp in length
117462 117561: gap of unknown length
117562 118402: contig of 841 bp in length
118403 118562: contig of 841 bp in length
```

```

* 118403 118502: gap of unknown length
* 118503 119341: contig of 839 bp in length
* 119342 119441: gap of unknown length
* 119442 120279: contig of 838 bp in length
* 120280 120379: gap of unknown length
* 120380 121210: contig of 831 bp in length
* 121211 121310: gap of unknown length
* 121311 122141: contig of 831 bp in length

Query Match
  Best Local Similarity 11.6%; Score 179.2; DB 39; Length 129226;
  Matches 262; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 518 tgcctgctggctgagtttgcctcctggtgacactgcccacccctgtggaaccaggct 577
Db 122325 TGGTTCCTGGAAAAATCTCCGGCAGGAGCTTAACRAAGTACATCCCAATATCTCGGCC 122384

QY 578 gcacagccctggccttcgcacaccacattgacctcaactgcagcgggcaagtgtccatct 637
Db 122385 TGGAGGCCATGGCCCTAAAGACCACCTATCGATCTTACCTGCAACGACTTCATTCCAACT 122444

QY 638 tcgagttcgagctttccaccaggctcttcagccatggcccaacactccctcaagaactggc 697
Db 122445 TCGAGTTTCGACGCTTTCACACGCTCTTCCAGCCTTGGGTGACCCCTGCTACGCAACTGGC 122504

QY 698 agctcctcgcagtcacacacccagggtacatggccttcctcactatgatgaggtccaaag 757
Db 122505 AGATTCTGGCCGTACACATCCGGGCTACGTGGCGTTTCTACATACGACGAGGTGAAGG 122564

QY 758 agcgtctgcagcctgcagggacaagccaggcagttacatcttccggccagctgtactc 817
Db 122565 CTGCGCTACAGCGCTACATCCTCAGCGGGGACGCTACGTTTCCGGGCTCTCTGACGCG 122624

QY 818 gcctggggcagtgggccatcgggtatgtcagctcagatggcagcatcctgcagaccatcc 877
Db 122625 GATTGGGCCAATGGCCATCGGCTACGTAACTGCCGAGGGAGAGATTCTGCAGACAATCC 122684

QY 878 ctgccacaacacccctgtccacagtgctcctcggaggagaca 917
Db 122685 CTCAGAACAAAGTCGCTGTGCCAGGCGCTGCTGATGGCCA 122724

```

Search completed: December 7, 2000, 13:43:21
 Job time: 27733 sec